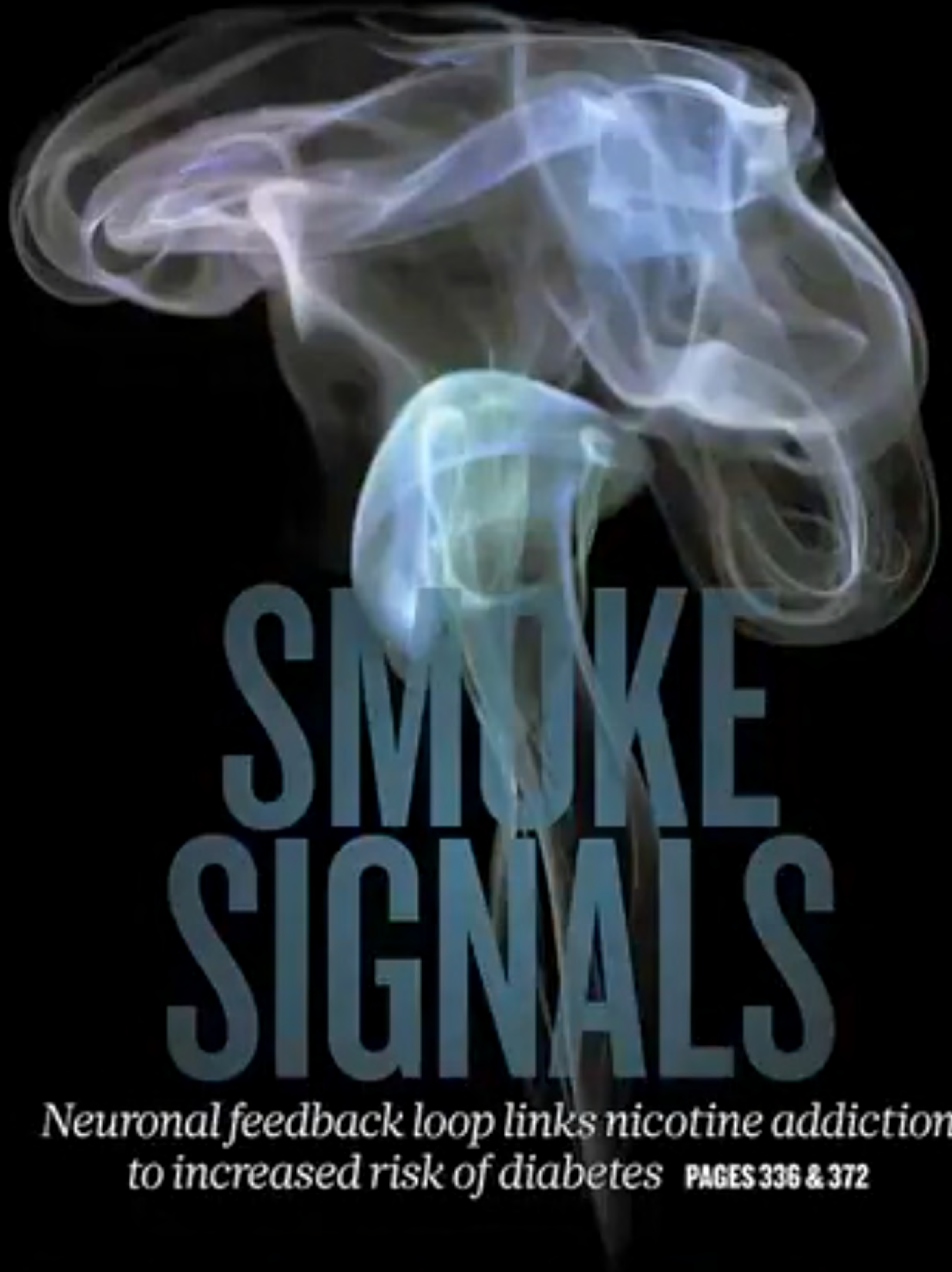
OUTLOOK
*RNA therapies*

# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



## SMOKE SIGNALS

*Neuronal feedback loop links nicotine addiction to increased risk of diabetes* **PAGES 336 & 372**

150 YEARS OF *NATURE*

### HOW DID DATA GET SO BIG?

*From empirical unit to social commodity*

**PAGE 317**

AUTUMN BOOKS

### SEASONAL READING

*Capitalism, pandemics and the state of the seas*

**PAGE 322**

HEALTH SCIENCE

### CHILD MORTALITY

*Map of 123 million deaths reveals regional inequalities*

**PAGE 353**

NATURE.COM
17 October 2019
Vol. 574, No. 7778

**WORLD VIEW** Childhood mortality data reveal realities of inequality **p.297**

**PREVIOUS CONVICTIONS** Shape-shifting cadmium-atom nucleus **p.299**

**TO BOLDLY GO** UK plans DARPA-style research funding agency **p.300**

# Boosting inclusivity in the Nobels

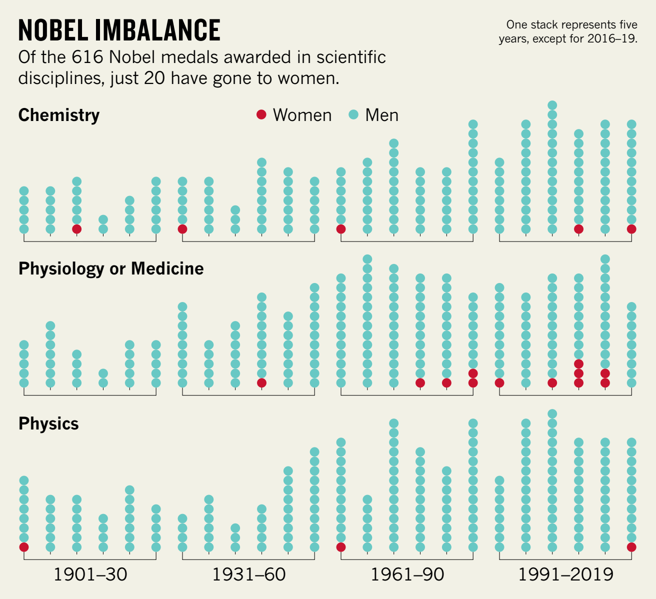
*After the euphoria of 2018, this year's Nobel prizes in chemistry, medicine and physics have again all been awarded to men. Here are three ways to encourage change.*

By Nobel standards, the committee that awards the economics prize is on a roll. Esther Duflo — awarded jointly for work that used randomized controlled trials to find ways to alleviate poverty — is the second woman to win the Nobel prize in economic sciences, following the late Elinor Ostrom in 2009. But, as we celebrate all of this year's inspiring Nobel prizewinners, we should reflect on the fact that, once again, members of the Royal Swedish Academy of Sciences and the Nobel Assembly of the Karolinska Institute have been unable to find a single recipient for the chemistry, physics and medicine awards who is not a man. Moreover, only one of the nine laureates in these categories — Japan's Akira Yoshino, awarded a share in the chemistry prize for his work on lithium-ion batteries — is not from Europe or North America.

The Royal Swedish Academy of Sciences is trying. More women are being nominated, the academy's secretary-general Göran Hansson told *Nature* (*Nature* http://doi.org/dcm9; 2019). And when individuals and institutions are asked to nominate potential Nobel candidates, they are advised to consider diversity in gender, geography and topic in their choices.

As this journal and many others have pointed out over the years, the lack of diversity among laureates is a systemic problem — and we accept it is not one that can be solved quickly. We know that today's prizes are often based on work carried out decades ago, when barriers in academia to women and other under-represented groups were even more formidable than they are today.

We also know that established institutions suffer from inertia and that selection processes have built-in biases, none of which can be overcome overnight. But even by the sometimes slow standards of change in science, the Nobel awarding bodies' steps towards reform are too incremental.



**NOBEL IMBALANCE**

Of the 616 Nobel medals awarded in scientific disciplines, just 20 have gone to women.

One stack represents five years, except for 2016–19.

We pointed out last year — and reiterate — that a first-order action is to release the data. Both the Swedish academy and the Karolinska Institute need to start reporting on the numbers of female scientists and scientists from other under-represented groups being nominated. We understand that convention prevents the identification of individuals for at least 50 years, but it should be possible to provide aggregate figures without identifying individuals. Reporting of these data would provide a greater incentive for change.

A second relatively simple action would be to diversify the sources that nominations — which are by invitation only — are accepted from. This month, many individuals and institutions will receive letters inviting nominations for the 2020 prizes. How many of these letters are sent beyond elite universities and academies isn't known.

As a small test case, *Nature* approached three of the world's largest international scientific networks that include academies of science in developing countries. They are the International Science Council, the World Academy of Sciences and the InterAcademy Partnership. Each was asked if they had been approached by the Nobel awarding bodies to recommend nominees for science Nobels. All three said no.

A third set of actions concerns ways in which the Nobel organizations can better understand their past. There's an evident injustice when more than a century of awards for the world's best scientists in chemistry, physics and medicine yields only 19 women (see 'Nobel imbalance') and not a single black person. If the Nobel archives were opened up to historians, more detail would come to light, allowing lessons to be learnt for the future. Several universities, including the Massachusetts Institute of Technology in Cambridge and the University of Glasgow, UK, have made their archives accessible in this way to take steps towards understanding their institutions' histories — in both cases, their links to slavery.

At the same time, both the Swedish academy and the Karolinska Institute must find ways to posthumously recognize scientists whose contributions to discovery and invention have been overlooked. Data released after the 50-year blackout lists the physicist Lise Meitner as having been nominated 48 times for either chemistry or physics without ever winning a Nobel.

At the end of his *Nature* interview, Hansson says that although it is important to have Nobel-prizewinning role models who are women and from diverse ethnicities, the prize will be devalued unless it goes to the most important discoveries and the individuals who made them. But these statements are not mutually exclusive.

As the world's children and young people absorb the news of this year's Nobels, many will be imagining that they, too, might make discoveries that could make them a future winner of the world's most prestigious science prizes. Unless the Nobel awarding bodies widen their search for nominations, they stand to miss out on many of these discoveries. ■

# Food audit

*Counting the hidden US$12-trillion cost of our dismal handling of global food resources.*

There's an unfolding tragedy at the heart of the world's food system and its cause lies mainly at the door of governments, food manufacturers and agribusinesses.

The situation is urgent. One-third of all food goes to waste, and yet governments and other players in the food system are unable to prevent 820 million people from regularly going hungry. The food industry, especially, bears responsibility for the fact that 680 million people are obese, but it is largely governments and their citizens who have to pick up the costs of treatment.

When industrial-scale farms draw copious quantities of water to irrigate crops, again it is taxpayers who foot the bill for the water scarcity that can follow. It's the same for agrochemicals and their effects on the health of people and ecosystems. Governments find themselves shouldering the costs of biodiversity loss, and mopping up agriculture's contribution to greenhouse-gas emissions.

These hidden costs — or externalities — must be met, and last month a landmark report estimated them to be somewhere in the region of US$12 trillion a year, rising to $16 trillion by 2050. That is a staggering figure — equivalent to the gross domestic product of China.

What is equally alarming is that these costs are not being regularly counted, and the food and agriculture industries seem to assume that the bill will be paid. That isn't right and has to change.

The report, which is the work of an organization called the Food and Land Use Coalition — which includes business groups and research institutions as well as the United Nations — also calculated the costs that governments and businesses would need to pay to transition to a more sustainable food system. That estimate comes to somewhere between $300 billion and $350 billion annually. In addition — and after taking account of hidden costs — a more sustainable food system could yield a further $5.7 trillion a year by 2030 in new economic opportunities, offsetting the $350-billion price tag by many multiples.

For example, a transition to plant-based diets containing less salt, sugar and processed foods is estimated to cost $30 billion. But the resulting economic benefits are predicted to be around $1.28 trillion. Cutting food waste is similarly estimated to cost $30 billion, with an estimated $455 billion expected to flow in commercial opportunities from waste reduction.

So if there's money to be made, it is reasonable to ask what is holding companies back. Why aren't they queueing up for a slice of the pie? Some undoubtedly are, but more could be persuaded, or compelled, to act.

> **"There must be more-regular accounting of these hidden costs."**

Governments have several levers when it comes to getting companies to change behaviour. One is taxation, a function of which is to fund public services, including clean-up efforts. Another lever is regulation — although in recent years, the fashion among some governments, in developed countries at least, has been to avoid imposing strong regulations. Instead, there is a move towards using softer methods to change practices in industry, drawing on the work of researchers in the behavioural sciences, for example.

A third lever is financial incentives — such as promoting the idea that companies can make profits from sustainability. Such an approach has had a measure of success following the influential 2006 publication of *The Economics of Climate Change: The Stern Review*, from development economist Nicholas Stern of the London School of Economics and Political Science. Among other things, this report and others that followed paved the way for several climate-change funding initiatives.

Whichever lever is used — and the most effective route is likely to involve a combination of all three, and more — there must be more-regular accounting and publishing of these hidden costs. That could be a task for national ministries of finance, or national statistics offices, working closely with researchers.

The Food and Land Use Coalition has performed an important service, but its calculations cannot be a one-off exercise, and governments, in turn, need to use these data to compel industry to act. ■

---

# Protect the census

*The census risks becoming a casualty of the rush to embrace big data. But it could save lives.*

Every child death is a tragedy. That's why a huge survey published this week on global child mortality is both striking and important. The work, by a collaborative team of some 4,000 people, is a detailed study of the variation in mortality rates among under-fives across 99 low- and middle-income countries between 2000 and 2017 (R. Burstein *et al. Nature* **574**, 353–358; 2019). It is one of the largest studies of its kind.

The work, led by Simon Hay at the University of Washington in Seattle and funded by the Bill & Melinda Gates Foundation, was carried out to help countries identify districts where under-fives are at greater risk of premature death.

Overall, deaths among under-fives have fallen from 19.6 million in 1950 to 5.4 million in 2017, thanks to steady gains in socio-economic development and public health. Today, the overwhelming majority of under-fives mortality is in low- and middle-income countries, which is why the researchers chose these countries to dig deeper.

The study is particularly valuable for its detail: it allows health-policy officials, researchers and non-governmental organizations to see district-level data. So although, at a national level, some countries — such as Colombia and Panama — have achieved the United Nations target of reducing infant mortality to fewer than 25 deaths per 1,000 live births by 2030, there are individual regions that are yet to reach the goal. Hay and his colleagues' work will help to identify these left-behind populations that, he says, "we need to pay more attention to" — with extra efforts to tackle poverty, inequality and poor health.

What is also remarkable about this piece of work is that it relied on a research method that risks going out of fashion. The team collected their information by poring over records of household surveys, including information on individual families that many countries record in a conventional census. More than 450 separate surveys provided the basis for the data, often asking the difficult questions of how many children a person has had — and how many have survived.

"We need more surveys," Hay told *Nature.* Unfortunately, in many countries, censuses are being scaled back. The conventional technique of processing paper-based forms, or sending interviewers to households, has always been expensive. That is one reason why, increasingly, households are being asked to fill out census data online — and it is possible that, at some point, this, too, will end.

After all, in many countries, swathes of digital data on citizens are already being collected online — from health and education records to election registration, tax returns and more. But these data may not be as comprehensive as a full-blown census, and in less-wealthy countries the data may not be collected at all, which is why censuses are still invaluable.

Hay and colleagues' work is a reminder that the conventional census has powerful uses. It is costly and time-consuming, but in this case it could demonstrably save lives. Governments should think harder before scaling back or abandoning censuses — and if they do, it's vital that other funders step in. ■


OHCHR

# Data on child deaths are a call for justice

*To save young lives, governments must support families, respect women and tackle inequality, says* **Michelle Bachelet***.*

The chances of newborns surviving to adulthood have never been greater: in the past 20 years, rates of childhood death have fallen by more than half. Nearly all 193 United Nations member states have made tremendous progress. But within each country, disparities condemn many children to premature death.

A study in *Nature* this week zooms in on the 99 low- and medium-income countries where, in 2017, 93% of deaths of children under 5 happened (R. Burstein *et al. Nature* **574**, 353–358; 2019). The authors estimate mortality for young children in each of 17,554 administrative regions from 2000 to 2017 — a remarkable level of detail.

The overall result is encouraging: 60% of the districts show sustained progress. But a closer look reveals continued inequalities. The research team calculates that 58% of the 123 million deaths mapped over 17 years were preventable. If all areas had mortality rates equal to the best-performing regions in their country, these deaths would not have occurred.

Richer areas, including capital cities, typically have much lower rates of child mortality. Poor governance can lead to the disproportionate allocation of public resources to elite groups. Deaths of small children spike in areas of poverty — where minority ethnic or religious groups, Indigenous peoples and other people who experience discrimination are likely to live. Discrimination against women leads to failure to prioritize maternal health; discrimination against ethnic or religious groups results in inadequate services for adults and children. And child deaths often accompany other human-rights violations. Insecurity, violence and conflict affect millions of children.

I COULD TRY TO CURE THEIR SICKNESS, BUT I COULD NOT PROVIDE A NOURISHING ENVIRONMENT.

The overarching lesson of these detailed mortality maps is that we must frame health programmes very broadly. Tackling child mortality requires efforts across all government functions — providing access to medical treatment is just one element. If governments are to uphold citizens' rights to health, they must consider the social determinants of children's well-being.

As a paediatrician treating the families of political prisoners in the 1980s during the dictatorship of Augusto Pinochet in Chile, I saw children who were not only ill, but also impoverished and traumatized. Back then, I could try to cure their sickness, but I could not provide a nourishing environment for them to grow up in. That is something I strove to do through policy in my work as Chile's minister for health and defence (2002–06) and president of Chile (2006–10 and 2014–18), and subsequently as a global health advocate, leader of UN Women and now UN high commissioner for human rights. Throughout my career, I have seen that improvements are feasible, cost effective and durably beneficial.

In Chile, death rates for children under 5 dropped from 148 to 7 per 1,000 between 1961 and 2018 — an unprecedented improvement. Since 1990, Chile's maternal mortality decreased from 31 to 13 per 100,000 live births, and infant mortality from 16 to 6 per 1,000. Disparities between districts also decreased, with the largest mortality reductions in the poorest areas.

How did we succeed? Our programmes focused on goals far beyond that of making sure that sick children could see a doctor.

As president of Chile, I knew that reducing child mortality required ensuring access to care, monitoring pregnant women's health and promoting vaccination campaigns. It also meant providing food security, in part through schools that offer nutritious meals. The 'Chile Grows With You' programme, launched in 2006, was the first of its kind in Latin America. It assists with offering education, maternity care and health services to families.

Providing adequate services is about more than resources. If women feel demeaned or inadequately consulted about their own bodies at hospital or physician visits, or if they are made to wait for hours at appointments, they might not turn up for prenatal care. We paid attention to differing needs across the country. We installed signage in Indigenous languages in key areas, introduced intercultural facilitators and encouraged inclusive approaches.

Truly universal and high-quality health coverage demands policies that extend beyond the strictly medical framework. It requires measures to uphold the whole range of human rights and to combat inequality and deprivation. Policies for decent housing, better labour rights, expanded childcare and unemployment benefits, access to basic services such as clean water — these are the types of reform that must be recognized as essential to health.

Childhood-mortality data reveal the stark realities of inequality. And improvements in these data mean fewer tragedies for families. They also shift the future of whole communities, potentially for generations. The World Bank has estimated that up to 30% of the differences in countries' per capita gross domestic product can be attributed to 'human capital' — essentially, education, training and health.

Three decades ago, the Convention on the Rights of the Child, which has become the most widely ratified human-rights treaty in history, emphasized the right of every newborn to the highest attainable standard of health. However complicated the ultimate causes, the factors that contribute to childhood deaths come down to failures to treat broader ills: poverty, disempowerment, discrimination and injustice.

Hard data, like those published this week, must be followed up by action across the whole spectrum of government and society. ■

---

**Michelle Bachelet** *is the UN High Commissioner for Human Rights and former two-term president of Chile.*
*Twitter: @UNHumanRights; @mbachelet*

# SEVEN DAYS *The news in brief*

CLIMATE CHANGE

## Climate lawsuit

Alaska's Supreme Court heard arguments on 9 October in a climate lawsuit that accuses the state government of violating the rights of young people by encouraging the use of fossil fuels. Sixteen children and young adults, ranging in age from 7 to 22, filed the case. They argue that the state of Alaska has a constitutional responsibility to protect the climate as a public resource for future generations. Their goal is to overturn a state law enacted in 2010 to promote fossil-fuel development. The supreme court will now determine whether the lawsuit, known as *Sinnok* v. *Alaska*, can go to trial. The case is one of several legal challenges filed by young environmentalists — and their lawyers — around the world. It is patterned after a similar lawsuit filed in 2011, which the Alaska Supreme Court dismissed in 2014 because it did not challenge specific actions by the government.

POLICY

## Misconduct law

China's science ministry released a draft on 8 October of its first law regulating research misconduct. The proposed law would make it illegal to: fail to declare a conflict of interest; embezzle research funds; breach intellectual-property rights; fabricate research results; commit plagiarism; and carry out research that "endangers national security". The law follows a series of government policies designed to crack down on problems that have plagued China's scientific community, such as data fabrication and faked authorship of research papers. Those caught committing misconduct could face penalties that include termination of research projects; loss of bonuses, awards and honorary titles; and being banned from applying for future government projects. Research managers can also be punished for covering up misconduct at their institution. The public can comment on the draft law until 8 November.

BEN SMEGELSKY/NASA

## Atmospheric-explorer mission lifts off

NASA launched a mission on 10 October to probe connections between weather on Earth and in space. The Ionospheric Connection Explorer (ICON) lifted off from Cape Canaveral Air Force Station, Florida, on a Pegasus XL rocket carried by an aeroplane; on separating from the plane, the rocket ignites its engines to reach orbit. ICON's launch had been delayed for more than a year by problems with the rocket. The satellite will study the links between the electrically neutral lower part of Earth's atmosphere and the electrically charged upper part, the ionosphere. Its four instruments will probe the region from about 90 kilometres to 580 kilometres above Earth's surface, measuring the ionosphere's density, structure and composition. Changes in the ionosphere can disrupt communications grids on Earth.

## Funding plans

The United Kingdom is set to get a new funding agency, broadly modelled on the US Defense Advanced Research Projects Agency, the British government announced on 14 October. The agency will bring a "new approach to funding emerging fields of research and technology", the government said in a document setting out its priorities at the formal opening of a new parliamentary session. Details on the value of the proposed fund have not yet emerged. Dominic Cummings, senior adviser to Prime Minister Boris Johnson, pitched the idea for such an agency on his blog, before taking up his current role. The government added that it would soon release spending plans that include a boost for research funding, and that it would act to reduce bureaucracy in research.

EVENTS

## Power blackout

California's largest utility company shut off power to more than one million people across the northern part of the state on 9 and 10 October. The outage sent scientists at the University of California, Berkeley, scrambling to save specimens and experiments. James Olzmann, a metabolic researcher at the university, moved his lab's freezers on 9 October to facilities at the University of California, San Francisco, which still had power. And Jessica Lyons, a molecular biologist at Berkeley, stocked freezers in her lab with dry ice on 8 October after being warned of the impending outages. The Pacific Gas and Electric Company, based in San Francisco, planned the outages to reduce the risk of wildfires, citing forecasts of strong, dry winds that have fanned devastating blazes across the region in recent years. The company has acknowledged that its equipment probably sparked California's deadliest known wildfire, which killed 86 people in 2018. It had restored power to most of its customers by 12 October.

## Climate protests

Hundreds of scientists worldwide are backing a campaign of civil disobedience in an attempt to force governments to act on climate change. "Continued governmental inaction over the climate and ecological crisis now justifies peaceful and non-violent protest and direct action, even if this goes beyond the bounds of the current law," reads a statement, which has been signed by more than 800 academics. Scientists have an obligation to protect the natural world, it says. About 20 scientists gathered in London on 12 October to read out the call to action (**pictured**). It comes as protesters from the campaign group Extinction Rebellion disrupt cities around the world. Police have arrested more than 1,400 people in connection with the protests in London alone.

HENRY NICHOLLS/REUTERS

**AWARDS**

## Nobel prizes

Three scientists whose research led to the development of lithium-ion batteries — which revolutionized energy storage — won the 2019 Nobel Prize in Chemistry on 9 October. The prize honoured the work, done in the 1970s and 1980s, of John Goodenough, Stanley Whittingham and Akira Yoshino (see page 308). On 14 October, the Nobel prize in economic sciences was awarded to three economists who used randomized controlled trials to determine how to lift people out of poverty. The winners are Michael Kremer at Harvard University in Cambridge, Massachusetts, and Abhijit Banerjee and Esther Duflo, both at the Massachusetts Institute of Technology, also in Cambridge. Duflo is only the second woman to win the economics Nobel.

**PEOPLE**

## Alexei Leonov dies

Soviet cosmonaut Alexei Leonov, the first person to walk in space, died on 11 October at the age of 85. In 1965, Leonov was orbiting Earth in the *Voskhod 2* spacecraft when he opened the hatch and exited the capsule for more than 12 minutes. His spacesuit puffed up and he had to vent some air before he could re-enter the spacecraft — an untested move hailed as a gutsy decision. Leonov later worked with US counterparts to help smooth the way for joint US–Soviet ventures in spaceflight during the cold war. He was also an accomplished artist, who created the first sketches of Earth from space.

**PUBLISHING**

## Open research

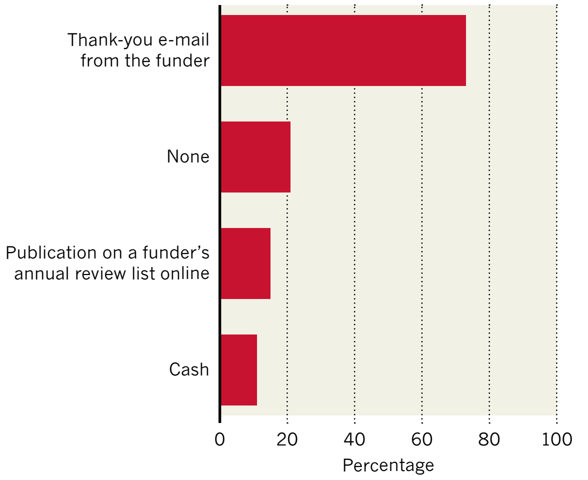
More than two-thirds of research papers read online will be open-access articles by 2025, according to a large-scale analysis of publishing and readership trends. The study, released on 9 October, used data from more than 50,000 journals and open-access repositories (H. Piwowar *et al.* Preprint at bioRxiv http://doi.org/dcm7; 2019). Of 70 million papers published between 1950 and 2019, fewer than one-third are currently freely available, it found. But more than half of papers presently viewed online are open-access articles. The study predicted, by extrapolating data from recent trends, that by 2025, articles published open access in one form or another will account for 70% of all papers viewed online, and that 44% of all papers published since 1950 will be free to read.

## TREND WATCH

More than half of researchers are more likely to accept requests to review grant applications if funders recognize their efforts, according to a survey of 4,700 researchers worldwide — the largest study of grant peer review so far. The *Grant Review In Focus* report, released on 10 October, was produced by Publons, a website that helps researchers to keep track of their peer-review contributions. It found that recognition is an important incentive for reviewers, who spend around ten days every year reviewing other academics' funding proposals. The process helps funders to make decisions, and although many researchers get acknowledgements such as thank-you e-mails, around 20% say they do not receive any recognition. Only a handful have received payment. However, the report suggests that greater financial rewards wouldn't necessarily sway reviewers into accepting more requests: cash came just sixth on the list of potential incentives. Instead, researchers prefer to have more recognition from employers and funders, to know the outcomes of the proposals they reviewed and to get feedback on the quality of their evaluation, the survey found.

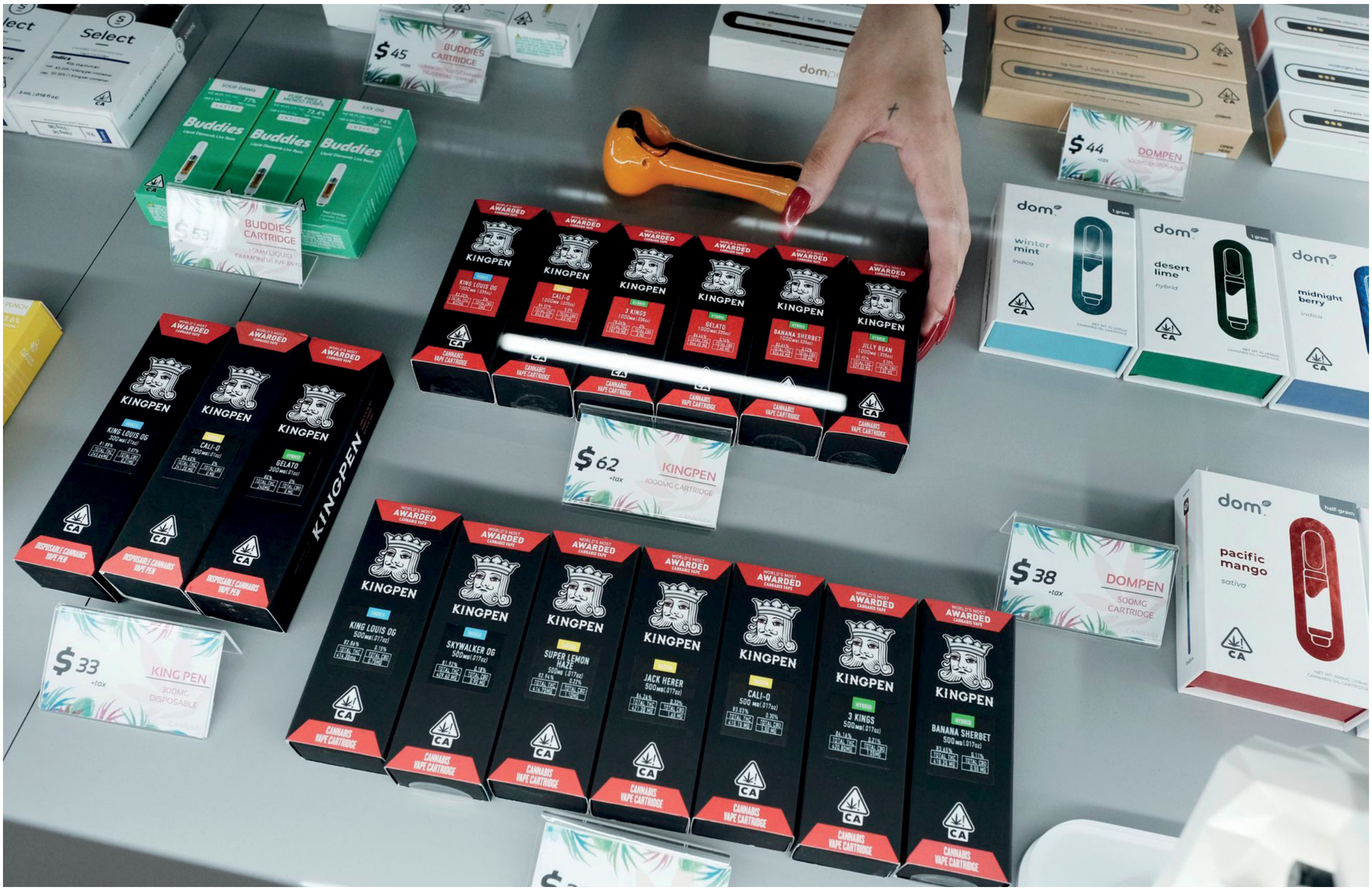
### REVIEWER REWARDS

Grant reviewers receive token recognition for their work, with fewer than one in five being paid.



SOURCE: GRANT REVIEW IN FOCUS

# NEWS IN FOCUS



RICHARD VOGEL/AP/SHUTTERSTOCK

The wide variety of vaping cartridges available for purchase has complicated efforts to trace the cause of a mysterious lung disease.

**PUBLIC HEALTH**

# Scientists chase cause of mysterious vaping illness

*Confusion reigns as researchers sort through the complex US e-cigarette market.*

**BY HEIDI LEDFORD**

Until a few months ago, pulmonologist Sean Callahan didn't typically ask his patients if they vaped. He thought that e-cigarettes might help smokers to wean themselves off cigarettes, and that the risks of vaping would probably take years to become clear.

The emergence of a mysterious, sometimes lethal, lung injury associated with vaping has changed his mind. Callahan works at the University of Utah Health in Salt Lake City, which has treated about 20 of those affected. "It was surprising: the overwhelming number of them — and how young they were," he says.

Researchers and physicians alike were caught unprepared by the illness, which has now sickened about 1,300 US vapers and killed 26. Scientists are scrambling to find out why, and to save other vapers from the same fate. "Everything is rapidly evolving," says Brandon Larsen, a pulmonary pathologist at the Mayo Clinic in Phoenix, Arizona. "I could tell you something today and next week it could be totally wrong."

A paper published by Larsen and his colleagues in the *New England Journal of* ▶

▶ *Medicine* on 2 October undercut a popular theory behind the outbreak — and underscored how far researchers still have to go to pinpoint its cause. (Y. M. Butt *et al. N. Engl. J. Med.* http://doi.org/gf9f5c; 2019).

Many of those who were made ill had vaped cartridges containing tetrahydrocannabinol (THC) — the active ingredient in marijuana — that had been diluted with oily chemicals.

Larsen's study is the largest analysis so far of lung tissue taken from affected vapers. The scientists searched for evidence of lipoid pneumonia, a condition that arises when oil enters the lungs. It is marked by the presence of lipids in lung tissue and in cells called macrophages, which normally sweep up debris in the lungs. But Larsen and his colleagues did not find substantial lipid droplets in any of their samples from 17 patients. Instead, their findings point to general lung damage and inflammation caused by exposure to toxic chemicals.

There are reasons to be sceptical of those results, says Kevin Davidson, a pulmonologist at WakeMed, a hospital system based in Raleigh, North Carolina. Larsen looked for signs of disease that would be apparent only if someone had inhaled a large amount of oil all at once, he says, not small amounts over time.

## DISEASE DETECTIVES

But Larsen's findings do align with mouse studies carried out by Farrah Kheradmand, a pulmonologist at Baylor College of Medicine in Houston, Texas (M. C. Madison *et al. J. Clin. Invest.* **129**, 4290–4304; 2019). Her team found lipids accumulating in the lung macrophages of mice exposed to e-cigarettes. The scientists traced the build-up to the breakdown of pulmonary surfactant, a lipid-rich compound produced by the lungs. Kheradmand says this suggests that vaping damages cells that line airways and help to maintain surfactant levels.

She is now hoping to repeat her mouse studies using e-cigarette vapour that contains vitamin E acetate, an oily chemical that has been suggested as a cause of the vaping illness.

Other researchers are considering similar experiments. Steven Rowe, a pulmonologist at the University of Alabama, Birmingham, hopes to test suspected culprits using ferrets, to learn how vaping affects ion transport in human lung cells. And Quan Lu, a lung biologist at the Harvard T. H. Chan School of Public Health in Boston, Massachusetts, is planning an experiment to see which genes are switched on or off in lung cells taken from vapers. He hopes to get tissue samples from those who have become ill.

> ***"I could tell you something today and next week it could be totally wrong."***

But Kheradmand cautions against hoping for quick answers: her initial mouse study took three and a half years to complete.

"Science will win at some point," says Albert Rizzo, chief medical officer at the American Lung Association in Chicago, Illinois. "But I don't think it's going to be as soon as people would like."

More immediately, researchers are scrambling to categorize the chemicals contained in e-cigarettes. That is no simple task when there are thousands of products available, and a culture of users modifying e-cigarettes and their contents to change characteristics such as flavour or amount of vapour produced. "This is a tough nut to crack, to be honest," says Larsen. "And that's where the research really needs to go: figuring out what the contents are in all of these things."

The range of chemicals that vapers are exposed to is dazzling, says Mignonne Guy, a biobehavioural researcher at Virginia Commonwealth University in Richmond. Her laboratory has studied YouTube videos and other online sources to learn more about how e-cigarette users are modifying their devices. They found that vapers are altering everything from how hot their e-cigarettes get to what chemicals are included in vaping cartridges — including, in at least one instance, liquid Viagra.

Online forums have pointed computational epidemiologist Yulin Hswen towards an early-2019 spike in posts about how to make e-cigarette cartridges. This was soon followed by an increase in posts from users warning about black-market cartridges being sold with branding that could mislead the purchaser into thinking they were made by a reputable company. Hswen, who works at the Harvard School of Public Health in Boston, Massachusetts, plans to look into this more closely to see whether this spike in homebrew cartridges could have contributed to the outbreak.

Ultimately, researchers might never be able to track down a single cause for the outbreak, says David Christiani, a pulmonologist at Harvard's public-health school. But even just narrowing it down to a process — such as using oils to dilute THC — could help to squelch the current epidemic and save lives. "We have a very serious epidemic and we absolutely need to get that under control," he says. "Then that will allow us to go back to focusing on chronic effects of vaping." ■

---

ARCHAEOLOGY

# Bronze Age DNA hints at roots of social inequality

*Family trees gleaned from ancient human genomes are set to transform archaeology.*

**BY EWEN CALLAWAY**

In a first-of-its-kind study, scientists have used DNA to reconstruct the family trees of dozens of individuals who lived in a small German valley around 4,000 years ago.

The genealogies point to social inequality within individual households, which encompassed both high-status family members and unrelated, low-status individuals — possibly servants or even slaves — as well as mysterious foreign females related to no one else (A. Mittnik *et al. Science* http://doi.org/gf9rmr; 2019).

Such insights could never have been made without using ancient DNA, says Philipp Stockhammer, an archaeologist at the Ludwig Maximilian University of Munich in Germany, who co-led the study. "For me, this is the future of archaeology," he says. "We are now forced to see social inequality and complexity on a completely different scale, that we haven't taken into account for the deep past."

During the Bronze Age, the Lech River Valley in southern Bavaria was packed with small farmsteads, each with its own cemetery. Archaeological excavations in the 1980s and 1990s uncovered dozens of skeletons dating to between about 2800 and 1700 BC.

Grave goods from these burials, such as daggers, arrowheads and ornaments, suggest that many Lech Valley inhabitants were well off, although the region lacks the mound-like 'princely graves' found elsewhere in Bronze Age Europe. Those often contain huge gold artefacts and show evidence of a social elite, archaeologists say.

K. MASSY

This ornate pin from the grave of a woman in Königsbrunn, Germany, suggests she had high social status.

To better understand the social structure of the Lech Valley, Stockhammer and Johannes Krause, at the Max Planck Institute for the Science of Human History (MPI-SHH) in Jena, Germany, and their team sequenced DNA from 104 individuals from 13 farmstead cemeteries. They constructed six family trees, which encompassed as many as five generations.

Nearly all first- and second-degree relationships the team uncovered were between individuals from the same farmstead; a few, more distant relations were found in different hamlets. These close family members, either male or female, tended to be buried alongside ample stashes of grave goods, suggesting high status was inherited. Cemeteries contained two other groups of individuals who were unrelated to any family members: people with poorly furnished graves, and high-status females.

It is impossible to say whether the low-status individuals represent servants, farm workers or slaves, according to the authors. But the social structures of the Lech Valley are reminiscent of those in ancient Greece and Rome, where slaves were considered members of the family unit.

The role of the high-status females is even more enigmatic. These women, who were buried with ornaments and jewellery similar to those of the female family members, grew up hundreds of kilometres away, Stockhammer says: the levels of strontium isotopes in their teeth, which vary with local geochemistry, are unlike those present in southern Germany. These women showed levels more similar to those found in eastern Germany and the Czech Republic.

But no children of theirs were found in the Lech Valley graves. One possibility is that females travelled hundreds of kilometres to the Lech Valley as part of alliances between wealthy families, and that any children were then returned to their mothers' native lands. The grave goods of some of the foreign females resemble those of the Únětice culture in the Czech Republic, Eastern Germany and Poland from around the same time.

## WHO'S WHO

"It's a really, really beautiful paper," says Kristian Kristiansen, an archaeologist at the University of Gothenburg, Sweden. "I know we'll see more of this." In unpublished work, Kristiansen and his colleagues sequenced DNA from more than 100 individuals from southern Germany and built family trees from the data.

"It does get to the heart of what archaeologists have been trying to do. They spend a lot of time working out who's related to who in cemeteries," says evolutionary geneticist Krishna Veeramah at Stony Brook University in New York. But he notes that few archaeologists will have access to the approach until sequencing costs come down.

The study marks a shift in how ancient genomics has been applied to archaeology, say Kristiansen and others. Many earlier studies — especially of Bronze Age Europe — sampled large numbers of unrelated individuals spread across dozens of sites in multiple countries. Many studies went on to document profound shifts in the genetic make-up of a region's inhabitants, to the chagrin of many archaeologists who tend to focus on local change and the lives of individuals.

"Instead of talking about a spread of an ancestry, we're really getting deep into the living history of these people," says Alissa Mittnik, a geneticist at MPI-SHH and Harvard Medical School in Boston, Massachusetts, who co-led the *Science* study. She hopes that shifts identified in earlier studies can be explained in more depth. For instance, her team reports that nearly all the women in the Lech Valley had moved away from their families — potentially spreading new cultural practices.

And as the number of sequenced ancient human genomes swells into the thousands, researchers will be able to build even bigger ancient family trees and identify distant relatives, just as customers of consumer-genetics firms such as 23andMe and Ancestry.com do today. Some of the individuals studied by Stockhammer, Mittnik and their collaborators turned out to be related to two other Lech Valley inhabitants whose genomes were sequenced as part of a 2015 study of 101 ancient humans (M. E. Allentoft *et al. Nature* **522**, 167–172; 2015). They were from a nearby burial, but, with luck, more distant connections will emerge, Mittnik says. "One day we'll find where these foreign women in the Lech Valley came from. That would be amazing." ■

ABK SÜD

Isotopic analysis revealed that this woman found near Kleinaitingen, Germany, grew up in a distant land.

ANNA SCHROLL/UIG/GETTY

**Protective suits are required for work in the highest-rated biosafety-level-4 labs.**

INFECTIOUS DISEASE

# Japan imports deadly viruses

*Ebola and other pathogens arrive ahead of the Olympics, but critics say the move is unnecessary.*

**BY MARK ZASTROW**

Japan is preparing for tens of thousands of international tourists to descend on Tokyo for the Olympic Games next year — and that includes being ready for unwanted biological visitors.

Last month, Japan imported Ebola and four other dangerous viruses in preparation for a possible outbreak at the event. The Japanese health ministry says that researchers will use the samples, which include Marburg virus, Lassa virus and the viruses that cause South American haemorrhagic fever and Crimean–Congo haemorrhagic fever, to validate tests under development.

The viruses' arrival represents the first time that pathogens rated biosafety-level-4 (BSL-4) — the most dangerous rating — have been allowed to enter the Japanese National Institute of Infectious Diseases (NIID), the only facility in the country operating at that level.

Japan's medical-science community welcomes the move. Although infectious-disease scientists say that the risk of an outbreak during the Olympics isn't much higher than at any other time, access to the live viruses will boost the country's capacity to handle infectious diseases in general — and to prepare for a bioterror attack.

Although the NIID's laboratory in Musashimurayama, Tokyo, was built to BSL-4 specifications in 1981, it operated as a BSL-3 lab for decades because of opposition from residents. In 2015, the health ministry and Musashimurayama's mayor agreed that it could operate as a BSL-4 lab, but the decision to import the five viruses was finalized only in July.

Japan's ability to study the most dangerous pathogens has lagged behind that of other advanced nations — both the United States and Europe have more than a dozen BSL-4 labs in operation or under construction, and China is building a network of at least five BSL-4 labs, with one already operational in Wuhan.

> *“An Ebola virus infection during the Olympics could have devastating consequences.”*

“This is a landmark time, a landmark event” for NIID, says Masayuki Saijo, director of the NIID department that is responsible for haemorrhagic-fever viruses.

But not everyone is pleased about the imported viruses. Some local residents have told Japanese media that scientists and the government are using the Olympics as a pretext to import the pathogens. And Richard Ebright, a molecular biologist and biosecurity specialist at Rutgers University in Piscataway, New Jersey, says that BSL-4 labs can be prepared to handle outbreaks of hazardous agents without the need to bring them into the country ahead of time. Storing dangerous viruses, even in a highly secure lab, increases the risk of an accidental or deliberate release, he says.

## WHAT'S THE RISK?

The NIID will use the live samples to validate tests it has developed to assess whether a person with one of the viruses is still infectious, says Saijo. The tests measure whether the person is generating antibodies that are capable of neutralizing the virus in question, which would suggest that they are recovering, and not infectious, he says. If there is a person with one of these viruses at the games, such a test could provide valuable information for assessing whether they can be discharged from hospital, he says.

The development of these tests will boost Japan's preparedness for such an event or a bioterror attack, says Saijo. Other Olympic host nations didn't have to specially import these viruses ahead of the games because they already had the pathogens in BSL-4 labs. The NIID will also continue developing more sensitive and accurate tests after the games. Saijo says that he understands opposition from local residents, but that the live viruses give Japanese researchers an important advantage in preparing against infectious diseases.

Elke Mühlberger, a microbiologist at Boston University in Massachusetts, thinks that a major outbreak of Ebola at the Olympics is unlikely because the infection is not transmitted through the air. But she says that Japan's plan to assess the NIID's tests with live viruses before the games makes sense, especially given the ongoing Ebola outbreak in the Democratic Republic of the Congo. “A report of an Ebola virus infection during the Olympics could have devastating consequences if the emergency responses were not professional,” she says.

But Mühlberger is sceptical about the usefulness of neutralizing-antibody tests to evaluate whether a patient can be released. She says the easiest way to determine whether someone is virus-free is to look at the amount of viral RNA in their body fluids. “I don't believe anybody would release a patient just because they have developed neutralizing antibodies,” she says.

## ANIMAL RESEARCH

Now that the NIID is allowed to handle BSL-4 pathogens, researchers there will also be able to study other dangerous viruses that might emerge in the region, says Mühlberger. The latest genome-sequencing technologies are revealing that Ebola-like viruses are more common than previously thought, she says. Three in the same family were discovered in animals in the past year: the Mengla virus in Chinese bats and two Ebola-like viruses found in fish in the East China Sea. "It is amazing how many animals are infected with viruses which are very closely related to very, very dangerous pathogens," she says.

It remains unknown whether these viruses can infect or harm humans, says Mühlberger. But their diversity is "pretty scary", she says. "These viruses are everywhere."

Virologist Ayato Takada at Hokkaido University in Sapporo, Japan, is also excited about being able to study BSL-4 pathogens in animals in Japan. Until now, researchers had to apply for access to BSL-4 labs overseas, which are in high demand. ■

---

GENOME EDITING

# Errors in CRISPR–baby study

*Geneticists retract paper that suggested first gene-edited babies might die early.*

**BY EWEN CALLAWAY**

A study that raised questions over the future health of the world's first gene-edited babies has been retracted because of key errors that undermined its conclusion.

The research, published in June in *Nature Medicine*[1], had suggested that people with two copies of a natural genetic mutation that confers HIV resistance are at an increased risk of dying earlier than other people. It was conducted in the wake of controversial experiments by the Chinese scientist He Jiankui, who had attempted to recreate the effects of this mutation in the gene *CCR5* by using the CRISPR gene-editing technique in human embryos. The twin girls born last year as a result of the work did not end up carrying this exact mutation, but the research attracted attention because of its potential relevance to such experiments.

However, studies[2–4] that looked anew at the *Nature Medicine* research find no evidence that people with the mutation die early. The erroneous conclusion about *CCR5* was caused by errors in how the mutation was identified in a population-health database.

"I feel I have a responsibility to put the record straight," says Rasmus Nielsen, a population geneticist at the University of California, Berkeley, who led the study, which the authors retracted on 8 October. Nielsen also co-authored one of the papers rebutting its findings.

Some researchers stress that because the twins did not receive exactly the same mutation that occurs naturally, the original research and its retraction would not necessarily offer insights into their health anyway. But the episode raises questions about how best to assess the safety of future attempts to edit genes in human embryos.

He Jiankui shocked the scientific world when he announced, in November 2018, that his team had used CRISPR to disable the *CCR5* gene in two babies born that month. He, who was at the time a biophysicist at the Southern University of China in Shenzhen, said he targeted *CCR5* because people with a 32-DNA-letter deletion known as delta-32 in the gene are resistant to HIV but seem not to experience significant related health problems.

He has not published data supporting his work, but his announcement — presented at a scientific meeting — indicated that, for one twin, both copies of *CCR5* were altered, whereas the other twin carried edits in just one of her two copies. None of the changes exactly matched the delta-32 variation.

## RESULTS NOT REPLICATED

Research has hinted that the delta-32 mutation, which is relatively common in people of European ancestry, might carry downsides — one small study[5] found that carriers were more likely than other people to die from influenza. To tackle the question in larger data sets, Nielsen and his Berkeley colleague Xinzhu Wei looked at the UK Biobank database, which contains genome and health data from 500,000 people.


A 32-letter deletion in the gene ***CCR5*** confers resistance to HIV.

GETTY

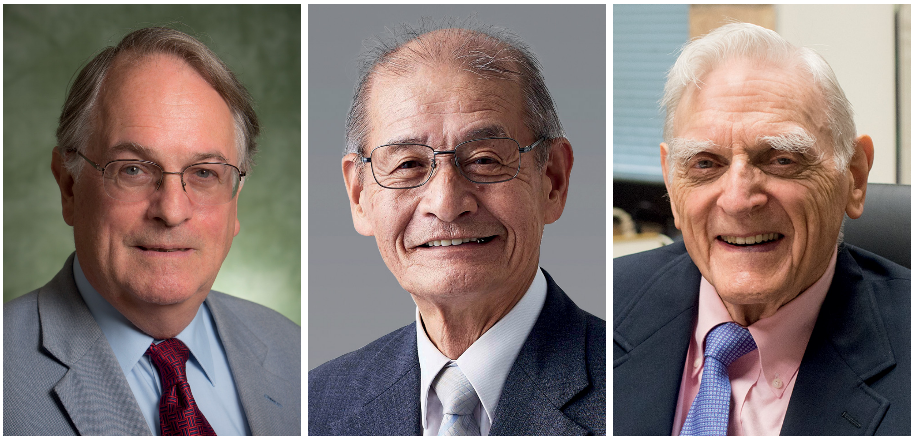
Their paper[1] reported that people with two copies of the delta-32 mutation were slightly more likely to die by the age of 76 than were those with one or no copies. They also found that the database harboured fewer people with two copies of the mutation than evolutionary theory predicted it should — a sign that individuals with two copies were dying earlier, on average, than the population at large, Wei and Nielsen argued.

Questions over the conclusion emerged as soon as the paper was published. Sean Harrison, an epidemiologist at the University of Bristol, UK, attempted to replicate the findings that night. He did not have UK Biobank data on the gene variant that Wei and Nielsen used to identify carriers of delta-32, so he analysed genetic variants near it on the genome that should have given the same result (adjacent parts of the genome tend to be inherited together, allowing scientists to infer the presence or absence of a DNA sequence by analysing neighbouring variants). When they didn't, he described his findings in tweets and a blogpost.

The discrepancy piqued the interest of David Reich, a population geneticist at Harvard Medical School in Boston, Massachusetts, whose lab is studying *CCR5*. Working with Nielsen, his team discovered[2] that Nielsen and Wei's method had caused them to undercount the number of people in the UK Biobank with two copies of the delta-32 mutation, because the probe that measured the variant they were tracking did not always identify its target sequence. This — and not the supposed harmful effects of the mutation — explained the apparent absence of carriers from the UK Biobank database, says Nielsen.

Researchers emphasize that the unravelling of Wei and Nielsen's results does not mean it is a sound idea to target *CCR5* for gene editing. "It's very reasonable to expect that it might have a valuable function that we just don't know how to measure. It seems very unwise to edit it out," says Reich. ■

1. Wei, X. & Nielsen, R. *Nature Med.* **25**, 909–910 (2019).
2. Maier, R. et al. Preprint at bioRxiv https://doi.org/10.1101/787986 (2019).
3. Gudbjartsson, D. *et al.* Preprint at bioRxiv https://doi.org/10.1101/788117 (2019).
4. Karczewski, K. J., Gauthier, L. D. & Daly, M. J. Preprint at bioRxiv https://doi.org/10.1101/784157 (2019).
5. Falcon, A. *et al. J. Gen. Virol.* **96**, 2074–2078 (2015).



Stanley Whittingham (left), Akira Yoshino and John Goodenough (right) share the Nobel chemistry prize.

L TO R: BINGHAMTON UNIV./THE JAPAN PRIZE/UNIV. TEXAS AT AUSTIN

CHEMISTRY PRIZE

# World-changing batteries win Nobel

*Prize honours three researchers who developed rechargeable lithium-ion batteries.*

BY DAVIDE CASTELVECCHI & EMMA STOYE

Three scientists whose work led to the development of lithium-ion batteries — a technology that ushered in a revolution in energy storage — won the 2019 Nobel Prize in Chemistry on 9 October.

John Goodenough, Stanley Whittingham and Akira Yoshino all contributed to the evolution of the kind of lightweight, rechargeable batteries that power today's mobile phones and other portable electronic devices. The technology also makes "possible a fossil fuel-free society", the Nobel Committee for Chemistry said.

Goodenough, a solid-state physicist at the University of Texas at Austin who is 97, became the oldest ever Nobel laureate. "I am very grateful that I had that honour, it's splendid," Goodenough told *Nature*. "But I am the same person I was before."

"Amazing! Surprising!" Yoshino, a chemist at Meijo University in Nagoya, Japan, told journalists after the announcement. The three researchers will get equal shares in the prize, which is worth 9 million Swedish kronor (US$910,000).

## RECHARGEABLE WORLD

"In my view, this award is long overdue," says Saiful Islam, a materials chemist at the University of Bath, UK. "It's great to see that this important area of materials chemistry has been recognized. As we know, these batteries have helped power the portable revolution."

In a lithium-ion battery, lithium ions move from the negative electrode (anode) to the positive electrode (cathode) through an electrolyte as the battery discharges, then travel back the other way when it is recharged.

While working for the oil company Exxon in the 1970s, Whittingham, who is now at the State University of New York at Binghamton, proposed the idea of rechargeable lithium batteries and developed a prototype that used a lithium-metal anode and a titanium disulfide cathode. The battery had a high energy density and the diffusion of lithium ions into the cathode was reversible, making the battery rechargeable. But high manufacturing costs and safety concerns meant the technology could not be commercialized.

In the late 1970s and early 1980s, Goodenough developed rechargeable batteries with cathodes made from layered oxides capable of storing lithium ions. This greatly improved the batteries' energy density, and lithium cobalt oxide remains the cathode material of choice for lithium-ion batteries.

And in the 1980s, Yoshino made changes to the materials that drastically improved safety and enabled commercial production of the batteries. His design pioneered the use of carbon-rich anode materials into which lithium ions could be inserted. "I started in 1981, and I invented the lithium-ion battery in 1985," Yoshino said.

Laboratories around the world are busy experimenting with technologies that could replace lithium-ion batteries, as well as with further developments to make existing batteries safer, more sustainable or longer lasting. But the fundamental design of the current battery will dominate for the foreseeable future, and — through electric vehicles and grid-scale energy storage — could play a crucial part in the decarbonization of the economy, says Clare Grey, a materials chemist at the University of Cambridge, UK. "It's poised to be around for a long time."

## SUSTAINABLE FUTURE

The Nobel committee also highlighted the role that the battery might have in creating a more sustainable future as nations try to move away from fossil fuels. Batteries are increasingly being used to store energy from renewable sources, such as solar and wind power, it noted.

"I salute the Nobel committee for recognizing something that is of practical importance," says Donald Sadoway, a chemist at the Massachusetts Institute of Technology in Cambridge. Islam adds that rechargeable batteries being developed for electric vehicles will have a crucial role in improving air quality, as well as lowering greenhouse-gas emissions.

Olof Ramström, a chemist at the University of Massachusetts Lowell who was a member of the Nobel chemistry committee, emphasized that this year's award recognizes a truly interdisciplinary effort. "It involved many branches of chemistry [and] it's also connected to physics and engineering," he said while announcing the prize. "It's a very good example of when all these disciplines can come together."

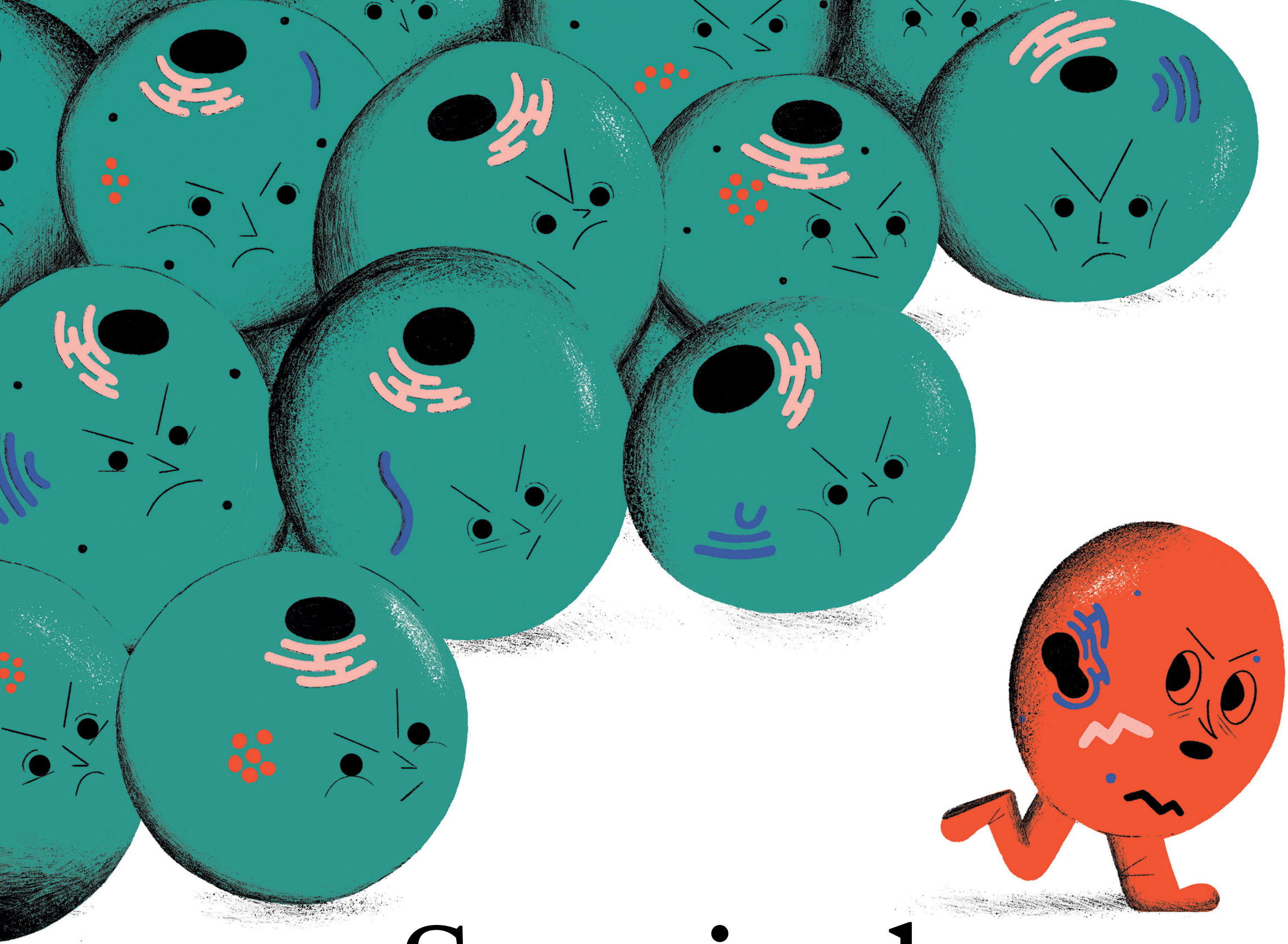
Grey, who has collaborated with Whittingham, says that a number of people have made major contributions to developing the technology, but that the committee made a good choice. Goodenough, in particular, is an "intellectual giant" in the field of materials research, and made many fundamental contributions beyond batteries. "He changed the way we think about magnetism, and he has helped to explain electronic conductivity," Grey says.

Asked whether lithium batteries were his favourite work, Goodenough told *Nature*: "No, I think my favourite work was to investigate what's called the Mott transition," a process in which the electrons in a material go from being able to move freely to being linked to individual atoms. ■

## CORRECTION

The News Feature 'Humans versus Earth' (*Nature* **572**, 168–170; 2019) incorrectly located the groups doing some work on samples from Crawford Lake. The testate amoebae work was done by a group at Carleton University, and the microplastics work was done at Brock University.

The News Feature 'The crystal kings' (*Nature* **572**, 429–432; 2019) erred in saying that Pablo Jarillo-Herrero was a graduate student of Philip Kim. He was a postdoctoral fellow.



# Survival of the fittest cells

Cells in the body don't always play nicely together. Could co-opting their competitive nature help to unlock cutting-edge therapies?

**BY KENDALL POWELL**

Yasuyaki Fujita has seen first-hand what happens when cells stop being polite and start getting real. He caught a glimpse of this harsh microscopic world when he switched on a cancer-causing gene called *Ras* in a few kidney cells in a dish. He expected to see the cancerous cells expanding and forming the beginnings of tumours among their neighbours. Instead, the neat, orderly neighbours armed themselves with filament proteins and started "poking, poking, poking", says Fujita, a cancer biologist at Hokkaido University in Sapporo, Japan. "The transformed cells were eliminated from the society of normal cells," he says, literally pushed out by the cells next door.

In the past two decades, an explosion of similar discoveries has revealed squabbles, fights and all-out wars playing out on the cellular level. Known as cell competition, it works a bit like natural selection between species, in that fitter cells win out over their less-fit neighbours. The phenomenon can act as quality control during an organism's development, as a defence against precancerous cells and as a key part of maintaining organs such as the skin, intestine and heart. Cells use a variety of ways to eliminate their rivals, from kicking them out of a tissue to inducing cell suicide or even engulfing them and cannibalizing their components. The observations reveal that the development and maintenance of tissues are much more chaotic processes than previously thought. "This is a radical departure from development as a preprogrammed set of rules that run like clockwork," says Thomas Zwaka, a stem-cell biologist at the Icahn School of Medicine at Mount Sinai in New York City.

But questions abound as to how individual

ILLUSTRATION BY ACAPULCO STUDIO

> “This is a radical departure from development as a preprogrammed set of rules that run like clockwork.”

cells recognize and act on weaknesses in their neighbours. Labs have been diligently hunting for — and squabbling over — the potential markers for fitness and how they trigger competitive behaviours. These mechanisms could allow scientists to rein in the process or to help it along, which might lead to better methods for fighting cancer and combating disease and ageing using regenerative medicine.

“Cell competition is on the global scientific map,” says Eugenia Piddini, a cell biologist at the University of Bristol, UK, who likens the buzz around this idea to the excitement that helped propel modern cancer immunotherapies. The better scientists understand competition, she says, the more likely it is that they will be able to use it therapeutically.

### HISTORY REPEATS

During a blizzard that dumped more than 30 centimetres of snow this past February, biologists from about a dozen disciplines convened at a hotel at Lake Tahoe, California, for the first major meeting devoted to cell competition.

“It was a zoo of researchers,” says co-organizer Zwaka, and included biologists who study flatworms that can regenerate their whole body from a single cell, geneticists attempting to make interspecies chimaeras of mouse, monkey and rabbit embryos, and a keynote speaker who spoke about the terrible battles and cooperative campaigns waged in bacterial communities.

The snowbound attendees, about 150 in all, debated how and why cells size up their competition. And they celebrated the discovery that gave birth to the field.

In 1973, two PhD students, Ginés Morata and Pedro Ripoll were perfecting a way to track the various cell populations in a fruit-fly larva that would eventually develop into a wing. Working at the Spanish National Research Council’s Biological Research Center in Madrid, they introduced a mutation called *Minute* into a few select cells in the larva and left the rest of the cells unaltered.

Knowing that *Minute* cells grow slower than their unaltered neighbours, the scientists expected to find some smaller cells amid the wild-type counterparts. “Instead, we found that the cells disappeared,” says Morata, now a developmental biologist at the Autonomous University of Madrid in Spain.

On their own, *Minute* cells can develop into a fly that is normal — except for the short, thin bristles on their bodies that give the mutation its name. But when mixed with wild-type cells in the larva, the cells simply vanished. “*Minute* cells were not able to compete with the more vigorous, metabolically active wild-type cells,” says Morata. They described the activity as cell competition[1]. “It was a very surprising and interesting observation,” Morata says. But lacking the molecular tools to follow cell fates more closely, he and his colleagues let the finding simmer.

Twenty-six years later, postdocs Laura Johnston and Peter Gallant observed nearly the same phenomenon. Working with Bruce Edgar and Robert Eisenman, respectively, at the Fred Hutchinson Cancer Center in Seattle, Washington, they were studying a mutation in another fly gene, *Drosophila Myc* (*dMyc*), that also slows cell growth[2].

“There was a eureka moment when Peter and I realized that these *dMyc* mutant cells would disappear,” says Johnston, now a developmental biologist at Columbia University Medical Center in New York City. They eventually showed that the mutant cells were forced to initiate a form of programmed cell death called apoptosis. “It was very clear that this was a competitive situation,” Johnston says.

Their 1999 paper ignited interest among scientists, including Morata. He jumped back into the fray with Eduardo Moreno, and they took advantage of modern molecular tools to repeat the *Minute* experiments. “The field blossomed from there,” says Johnston.

*Myc* acts as a master controller of cell growth, and *Minute* encodes a key component needed for synthesizing proteins — so it’s not surprising that reduced expression of those proteins makes cells less fit. But Johnston’s next finding took people by surprise. She showed that cells with an extra copy of normal *dMyc* outcompeted wild-type cells[3]. These fitter-than-wild-type cells came to be called “supercompetitors”.

Johnston’s discovery of supercompetition emphasized that cell competition is about the relative fitness of a group of cells, says Zwaka. If one cell is falling behind, the entire group of neighbours could decide it has to go. But on the flipside, they can also sense that certain cells are better and should survive.

Cell competition wasn’t simply about getting rid of defects; it was about survival of the fittest, with the less-fit ‘loser’ cells dying and the ‘winners’ proliferating. Importantly, competition was seen only when there was a mixture of genetically different cells, a phenomenon known as mosaicism. In this way, cell competition acts like a quality-control system, booting out undesirable cells during development.

### VYING FOR VIABILITY

Fujita’s observation of the kicked-out kidney cells was one of the first hints that mammalian cells compete, too[4]. Soon after that work was published, researchers started to observe competition forcing out mutated cells from various other tissue types such as skin, muscle and gut.

The next most obvious place to look for competing cells was the mammalian embryo. In 2013, Zwaka’s team, and two other laboratories, probed mouse embryos at the earliest stage of development — those that have progressed just beyond a ball of cells. Zwaka’s group made mouse embryonic stem cells (ESCs) with a supercompetitor mutation that lowered expression of p53, an important quality-control protein that normally puts the brakes on cell division. When these cells were put into a mouse embryo, they quickly took over and developed into a normal mouse[5]. Similarly, Miguel Torres’s lab at the National Center for Cardiovascular Research in Madrid showed that supercompetition could be induced in an early mouse embryo using slight overexpression of the mouse *Myc* gene.

By artificially creating losers or winners, researchers could force cell competition into play. But Torres’s team, led by then-postdoc Cristina Clavería, also made the striking observation that *Myc* expression varied naturally in mouse ESCs. Cells in the embryo with approximately half the amount of the protein compared with their neighbours were dying by apoptosis. This was one of the first studies that strongly pointed to naturally arising cell competition[6].

### SCULPTING TISSUES

The phenomenon also comes into play later on in embryonic development. In a study published this year, postdoc Stephanie Ellis at Elaine Fuch’s lab in Rockefeller University in New York City, looked at mouse skin. During development, its surface area expands by a factor of 30 over the course of about a week. The cells within proliferate wildly — first as a single layer and later as multiple layers.

Ellis injected mouse embryos with a concoction that turns cells into genetic losers. She targeted a few cells present when the embryonic skin is a single layer thick, and added a marker gene that made them glow red. Then she used time-lapse imaging to watch their grim fates: the skin cells popped out from the surface layer, broke up and disappeared. Later, she noticed the winner cells engulfing and clearing the losers’ corpses[7].

Repeating the experiment at the multilayer stage, Ellis no longer saw the less-fit skin cells perishing or being engulfed. Instead, the loser cells tended to differentiate and migrate into the outer layers of skin — where they acted as a barrier for a short time before being shed. The winner cells were more likely to remain behind in the bottom layer as stem cells.

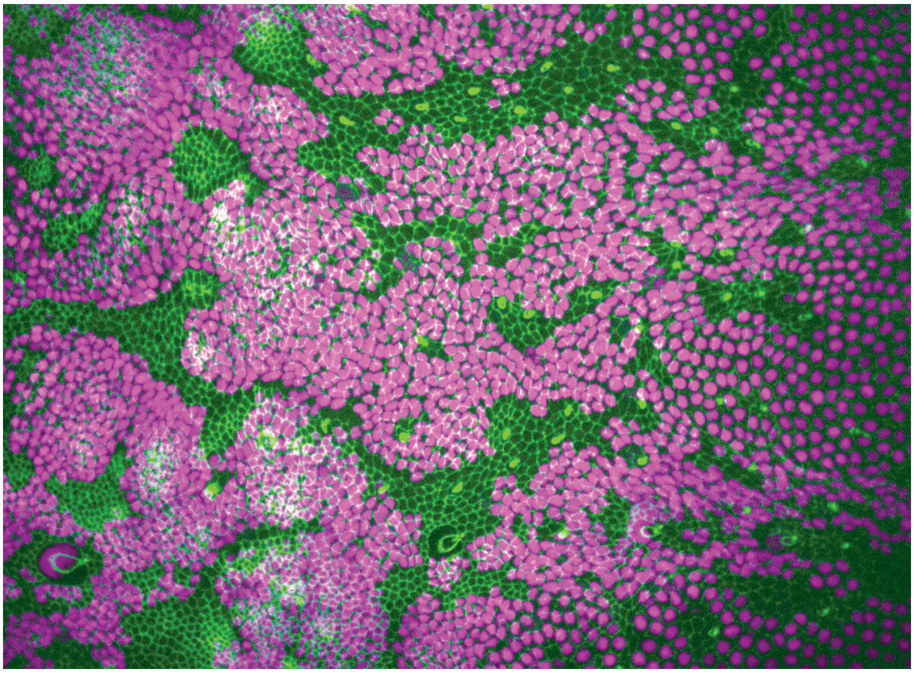
This made sense. “Killing a cell is energetically expensive,” says Ellis. A developing tissue, she says, might decide: “Why not just remove losers through differentiation?” Emi Nishimura’s lab at the Tokyo Medical and Dental University in Japan, found that competing stem cells in the ageing tail skin of adult mice used the same pattern of asymmetrical divisions to eliminate stem cells with lower levels of a key structural collagen protein[8].

These experiments could provide guidance for scientists looking to harness stem cells to rejuvenate ageing tissues and organs. Cell competition could either help or hurt such therapies: stem cells might outcompete older, less-fit cells, or they might encounter a hostile neighbourhood when transplanted into tissue. Understanding whether and how cell competition happens in adult tissue could help settle this matter.

Piddini admits that she was a little obsessed with the idea, and her group was part of a wave of researchers that proved cell competition does take place in adult organisms. To test the idea, she says, the team "genetically sprinkled" a mutated copy of *RPS3*, a gene functionally related to *Minute*, into some cells in the intestine of adult flies. Cells with the mutant copy were outcompeted by their wild-type counterparts. It didn't matter whether the losers were the stem cells that maintain the gut or differentiated cells: all eventually perished[9].

Cristina Villa del Campo, a senior postdoc in the Torres lab, tested for adult competition in the mouse heart by introducing winner cardiac cells at eight to ten weeks of age. Over the course of one year, she tracked the numbers of winner cells and wild-type losers and saw the loser population decline by about 40%[10].

"It was a slow replacement in the adult," says Villa del Campo. "But even highly differentiated functional adult cells can sense the less-fit heart cells and eliminate them."


ROMAIN LEVAYER, EDUARDO MORENO

**Supercompetitive pre-tumour cells (purple) in developing fly tissue squish and squeeze their wild-type counterparts (green).**

## UNANSWERED QUESTIONS

Even with so many examples of cell competition playing out in different conditions, the field still faces a torrent of unanswered questions. One big puzzle is how cells in a group sense fitness. "Maybe cells are recognizing chemical differences, or physical differences, or differences in cell-membrane composition," says Fujita, who adds that labs have found evidence for all three.

His filament-poking kidney-cell experiments suggest that cell–cell contact is needed. Others have seen chemical-fitness signals that seem to be short-range, travelling up to eight cell diameters. Exactly which molecules are responsible for this signalling — either secreted chemicals or physical tags — is the subject of intense debate and investigation.

Both Johnston and Zwaka have turned up signals associated with immune surveillance. Johnston's group identified molecules that typically call immune cells to swarm in and engulf foreign invaders and that were driving death in losers[11]. Normal cells express low levels of these death signals at all times. But in a competitive mix, winners flooded their loser neighbours with the signal, which pushed them to kill themselves.

Zwaka proposes that cells might assess each other's health by sniffing out the general signals or debris that cells shed. It's akin to smelling the steaks that your neighbour is grilling for dinner and concluding that they must be doing well.

Or it could be as simple as seeing which flag your neighbour is flying. Moreno heads his own group now at the Champalimaud Centre for the Unknown in Lisbon, Portugal, which discovered a membrane-spanning protein called Flower[12]. In humans, the protein can take four forms, each displaying its own characteristic structure on the outer cell surface. Two signal 'I'm a winner' and the other two signal 'I'm a loser' to nearby cells, says Moreno.

Some human cancer cells fly the Flower-winner signals, which might enhance their survival. Experiments in Moreno's lab showed that silencing the winning flags on tumours slowed the cells' growth and made them susceptible to chemotherapy[13].

Some researchers, however, dispute the importance of the Flower tags. Moreno acknowledges that they are not present in all cell-competition situations.

## HEALTHY COMPETITION

Cracking the mechanics of competition will be key if researchers want to use it to improve cell-based cancer or regenerative therapies.

There are tantalizing hints that cell competition might already protect against cancer. Findings made in the past few years reveal that human skin, oesophageal and lung cells show high levels of mosaicism. Approximately one-quarter of skin cells, for example, harbour many precancerous mutations that only rarely turn into tumours[14,15].

It is unclear what gives cancerous cells the advantage when tumours do form. If researchers can learn how to subdue supercompetitors or blunt cancer cells' ability to compete, they might be able to turn that against cancer.

Conversely, stem cells might need to gain a competitive edge if they are to replace aged or diseased tissue for an organ makeover. Villa del Campo says that clinicians are already considering how to enhance patient-derived cardiac stem cells to efficiently replace cells that have been damaged by heart attacks or disease.

What started as modest observations in minuscule fruit-fly larvae has exposed the primal cellular battles that could usher in a new era of cell-based medicine. The process has scientists buzzing, but it remains mysterious.

"Cell competition might be a general process to remove any undesirable cell that should not be there," says Morata, after returning from a one-day meeting in Lausanne, Switzerland devoted to competition in September. Now 74, he's thrilled that work he essentially shelved more than 40 years ago is gaining new life and that the competition is heating up. "It's really exciting." ■

---

**Kendall Powell** *is a science journalist in Boulder, Colorado.*

1. Morata, G. & Ripoll, P. *Develop. Biol.* **42**, 211–221 (1975).
2. Johnston, L. A., Prober, D. A., Edgar, B. A., Eisenman, R. N. & Gallant, P. *Cell* **98**, 779–790 (1999).
3. de la Cova, C., Abril, M., Bellosta, P., Gallant, P. & Johnston, L. A. *Cell* **117**,107–116 (2004).
4. Hogan, C. *et al. Nature Cell Biol.* **11**, 460–467 (2009).
5. Dejosez, M., Ura, H., Brandt, V. L. & Zwaka, T. P. *Science* **341**,1511–1514 (2013).
6. Clavería, C., Giovinazzo, G., Sierra, R. & Torres, M. *Nature* **500**, 39–44 (2013).
7. Ellis, S. J. *et al. Nature* **569**, 497–502 (2019).
8. Liu, N. *et al. Nature* **568**, 344–350 (2019).
9. Kolahgar, G. *et al. Dev. Cell* **34**, 297–309 (2015).
10. Villa Del Campo, C., Clavería, C., Sierra, R. & Torres, M. *Cell Rep.* **8**, 1741–1751 (2014).
11. Meyer, S. N. *et al. Science* **346**, 1258236 (2014).
12. Rhiner, C. *et al. Dev. Cell* **18**, 985–998 (2010).
13. Madan, E. *et al. Nature* **572**, 260–264 (2019).
14. Martincorena, I. *et al. Science* **348**, 880–886 (2015).
15. Yizhak, K. *et al. Science* **364**, eaaw0726 (2019).

PAUL NICKLEN/NG IMAGE COLLECTION

# GENES FROM THE JUNKYARD

*Scientists long assumed that new genes appear when evolution tinkers with old ones. It turns out that natural selection is much more creative.*

**BY ADAM LEVY**

In the depths of winter, water temperatures in the ice-covered Arctic Ocean can sink below zero. That's cold enough to freeze many fish, but the conditions don't trouble the cod. A protein in its blood and tissues binds to tiny ice crystals and stops them from growing.

Where codfish got this talent was a puzzle that evolutionary biologist Helle Tessand Baalsrud wanted to solve. She and her team at the University of Oslo searched the genomes of the Atlantic cod (*Gadus morhua*) and several of its closest relatives, thinking they would track down the cousins of the antifreeze gene. None showed up. Baalsrud, who at the time was a new parent, worried that her lack of sleep was causing her to miss something obvious.

But then she stumbled on studies suggesting that genes do not always evolve from existing ones, as biologists long supposed. Instead, some are fashioned from desolate stretches of the genome that do not code for any functional molecules. When she looked back at the fish genomes, she saw hints this might be the case: the antifreeze protein — essential to the cod's survival — had seemingly been built from scratch[1].

The cod is in good company. In the past five years, researchers have found numerous signs of these newly minted '*de novo*' genes in every lineage they have surveyed. These include model organisms such as fruit flies and mice, important crop plants and humans; some of the genes are expressed in brain and testicular tissue, others in various cancers.

*De novo* genes are even prompting a rethink of some portions of evolutionary theory. Conventional wisdom was that new genes

**Some cod species have a newly minted gene involved in preventing freezing.**

tended to arise when existing ones are accidentally duplicated, blended with others or broken up, but some researchers now think that *de novo* genes could be quite common: some studies suggest at least one-tenth of genes could be made in this way; others estimate that more genes could emerge *de novo* than from gene duplication. Their existence blurs the boundaries of what constitutes a gene, revealing that the starting material for some new genes is non-coding DNA (see 'Birth of a gene').

The ability of organisms to acquire new genes in this way is testament to evolution's "plasticity to make something seemingly impossible, possible", says Yong Zhang, a geneticist at the Chinese Academy of Sciences' Institute of Zoology in Beijing, who has studied the role of *de novo* genes in the human brain.

But researchers have yet to work out how to definitively identify a gene as being *de novo*, and questions still remain over exactly how — and how often — they are born. Scientists also wonder why evolution would bother making genes from scratch when so much gene-ready material already exists. Such basic questions are a sign of how young the field is. "You don't have to go back that many years before *de novo* gene evolution was dismissed," Baalsrud says.

## NEW ARRIVALS

Back in the 1970s, geneticists saw evolution as a rather conservative process. When Susumu Ohno laid out the hypothesis that most genes evolved through duplication[2], he wrote that "In a strict sense, nothing in evolution is created *de novo*. Each new gene must have arisen from an already existing gene."

Gene duplication occurs when errors in the DNA-replication process produce multiple instances of a gene. Over generations, the versions accrue mutations and diverge, so that they eventually encode different molecules, each with their own function. Since the 1970s, researchers have found a raft of other examples of how evolution tinkers with genes — existing genes can be broken up or 'laterally transferred' between species. All these processes have something in common: their main ingredient is existing code from a well-oiled molecular machine.

But genomes contain much more than just genes: in fact, only a few per cent of the human genome, for example, actually encodes genes. Alongside are substantial stretches of DNA — often labelled 'junk DNA' — that seem to lack any function. Some of these stretches share features with protein-coding genes without actually being genes themselves: for instance, they are littered with three-letter codons that could, in theory, tell the cell to translate the code into a protein.

It wasn't until the twenty-first century that scientists began to see hints that non-coding sections of DNA could lead to new functional codes for proteins. As genetic sequencing advanced to the point that researchers could compare entire genomes of close relatives, they began to find evidence that genes could disappear rather quickly during evolution. That made them wonder whether genes could just as quickly spring into being.

In 2006 and 2007, evolutionary geneticist David Begun at the University of California, Davis, published what many regard as the first papers to make the case for particular genes arising *de novo* in fruit flies[3,4]. The studies linked these genes to male reproduction: Begun found they were expressed in the testes and the seminal fluid gland, where it seemed the powerful evolutionary force of sexual selection was driving gene birth.

Shortly before that, evolutionary genomicist Mar Albà at the Hospital del Mar Medical Research Institute in Barcelona, Spain, had shown that the younger a gene is, evolutionarily speaking, the faster it tends to evolve[5]. She speculated that this might be because the molecules encoded by younger genes are less polished and need more tuning, and that this could be a consequence of the genes having arisen *de novo* — they were not tied to a previous function as tightly as those that had evolved from older genes. Both Albà and Begun recall that it was challenging to publish their early work on the topic. "There was a lot of scepticism," says Albà. "It's amazing how things have changed."

"IT'S LIKE A BETA LAUNCH."

Studies have also started to unpick what *de novo* genes do. One gene allows the thale cress plant (*Arabidopsis thaliana*) to produce starch, for instance, and another helps yeast cells to grow. Understanding what they are doing for their hosts should help to explain why they exist — why it is advantageous to create from scratch rather than evolve from existing material. "We're not going to understand why these genes are evolving if we don't understand what they're doing," says Begun.

## GENES-IN-WAITING

Studying *de novo* genes turns out to be part genetics, part thought experiment. "Why is our field so difficult?" asks Anne-Ruxandra Carvunis at the University of Pittsburgh in Pennsylvania. "It is because of philosophical issues." At its heart is a question that Carvunis has been asking for a decade: what is a gene?

A gene is commonly defined as a DNA or RNA sequence that codes for a functional molecule. The yeast genome, however, has hundreds of thousands of sequences, known as open reading frames (ORFs), that could theoretically be translated into proteins, but that geneticists assumed were either too short or looked too different from those in closely related organisms to have a probable function.

When Carvunis studied yeast ORFs for her PhD, she began to suspect that not all of these sections were lying dormant. In a study[6] published in 2012, she looked at whether these ORFs were being transcribed into RNA and translated into proteins — and, just like genes, many of them were — although it was unclear whether the proteins were useful to the yeast, or whether they were translated at high enough levels to serve a function. "So what is a gene? I don't know," Carvunis says. What she thinks she has found, though, is "raw material — a reservoir — for evolution".

Some of these genes-in-waiting, or what Carvunis and her colleagues called proto-genes, were more gene-like than others, with longer sequences and more of the instructions necessary for turning the DNA into proteins. The proto-genes could provide a fertile testing ground for evolution to convert non-coding material into true genes. "It's like a beta launch," suggests Aoife McLysaght, who works on molecular evolution at Trinity College Dublin.

Some researchers have gone beyond observation to manipulate organisms into expressing non-coding material. Michael Knopp and his colleagues at Uppsala University, Sweden, showed that inserting and expressing randomly generated ORFs into *Escherichia coli* could enhance the bacterium's resistance to antibiotics, with one sequence producing a peptide that increased resistance 48-fold[7]. Using a similar approach, Diethard Tautz and his team at the Max Planck Institute for Evolutionary Biology in Plön, Germany, showed that half of the sequences slowed the bacterium's growth, and one-quarter seemed to speed it up[8] — although that result is debated. Such studies suggest that peptides from random sequences can be surprisingly functional.

But random sequences of DNA could also code for peptides that are "reactive and nasty and have a tendency to aggregate and do bad things", says evolutionary biologist Joanna Masel of the University of Arizona in Tucson. Expressing these sequences at low levels could help natural selection to weed out potentially dangerous portions — those that create messy or misfolded proteins — so that what remains in a species is relatively benign.

Creating genes from non-coding regions could have some benefits over other gene-making methods, says Albà. Gene duplication is a "very conservative mechanism" she says, producing well-adapted proteins cut from the same cloth as their ancestors; *de novo* genes, by contrast, are likely to produce markedly different molecules. That could make it difficult for them to fit into well-established networks of genes and proteins — but they could also be better suited to certain new tasks.

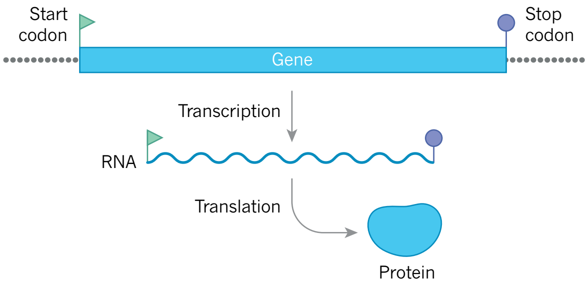
A newly minted gene could help an organism to respond to a change in its environment, for instance. This seems to have been the case for the cod, which acquired its antifreeze protein as the Northern Hemisphere cooled some 15 million years ago.

### BIRTH OF A GENE

Scientists long assumed that evolution made new genes from old ones — by copying them in error, or by fusing together or breaking apart existing ones. Now, more and more examples are emerging of genes being created '*de novo*', from barren non-coding portions of the genome.

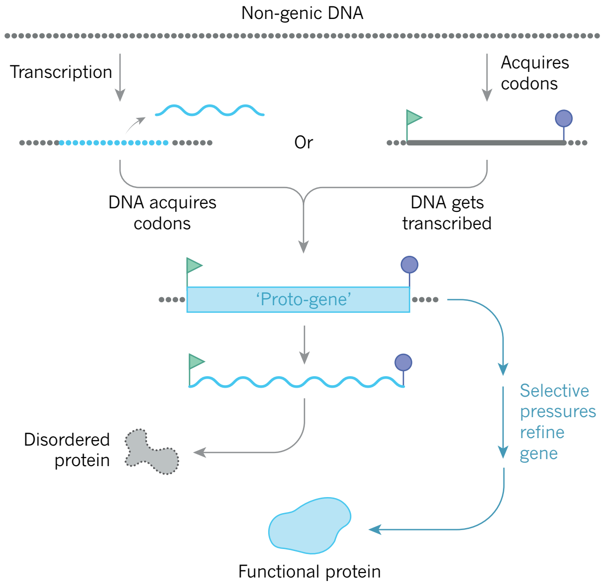
#### How genes work

Genes are usually considered any stretches of DNA that code for useful molecules. To make a protein, DNA is transcribed into RNA, which is then translated. Three-letter pieces of sequence called codons dictate which portions of the RNA to translate.



#### Making a gene *de novo*

Genes can evolve from non-coding portions of DNA by gaining transcription and codons, in either order. At first, the products of these 'proto-genes' might be dysfunctional or disordered.



NIK SPENCER/NATURE

### BIRTH RATE

To trace which of an organism's genes were made *de novo*, researchers need comprehensive sequences for the organism and its close relatives. One crop plant that fits the bill is rice. The sweltering heat of Hainan, a tropical island in southern China, is the perfect environment for growing the crop — although the working conditions can be trying. "It's horrible," says evolutionary geneticist Manyuan Long of the University of Chicago, Illinois. It's so hot "you can cook your egg in the sand".

Long's team wanted to know how many genes had emerged *de novo* in the strain *Oryza sativa japonica*, and what proteins those genes might be making. So the team lined up its genome against those of its close relatives and used an algorithm to pick out regions that contained a gene in some species but lacked it in others. This allowed the researchers to identify the non-coding DNA that led to the gene in question, and track its journey to being a gene. They could also tot up the number of *de novo* genes that appeared in the strain: 175 genes over 3.4 million years of evolution[9] (over the same period, the strain gained 8 times as many genes from duplication).

The study gets at one of the field's biggest preoccupations: how to tell whether a gene is truly *de novo*. Answers vary wildly, and approaches are still evolving. For example, an early study found 15 *de novo* genes in the whole primate order[10]; a later attempt found 60 in humans alone[11]. One option for finding candidate *de novo* genes is to use an algorithm to search for similar genes in related species. If nothing shows up, then it's possible that the gene arose *de novo*. But failing to find a relative doesn't mean no relative is there: the gene could have been lost along the way, or might have shape-shifted far away from its kin. The rice study got around this by explicitly identifying the pieces of non-coding DNA that became *de novo* genes.

Over long evolutionary timescales — much longer than the few million years of rice evolution — it is hard to distinguish between a *de novo* gene and one that has simply diverged too far from its ancestors to be recognizable, so determining the absolute number of genes that have arisen *de novo* rather than from duplication "is an almost unanswerable question", says Tautz.

To demonstrate how varied the results of different methods can be, evolutionary geneticist Claudio Casola at Texas A&M University in College Station used alternative approaches to reanalyse the results of previous studies, and failed to verify 40% of the *de novo* genes they had proposed[12]. To Casola, this points to the need to standardize tests. Currently, he says, "it seems to be very inconsistent".

Counting *de novo* genes in the human genome comes with the same trail of caveats. But where *de novo* genes have been identified, researchers are beginning to explore their roles in health and disease. Zhang and his colleagues have found that one gene unique to humans is expressed at a greater level in the brains of people with Alzheimer's disease[13], and previous work[14] had linked certain variants of the gene to nicotine dependence. For Zhang, research that links *de novo* genes to the human brain is tantalizing. "We know that what makes us human is our brains," he says, "so there must be some genetic kit to push the evolution of our brain." That suggests an avenue for future studies. Zhang suggests that researchers could investigate the genetic kit through experiments with human organoids — cultured cells that serve as a model organ.

*De novo* genes could have implications for understanding cancer, too. One such gene — unique to humans and chimpanzees — has been linked to cancer progression in mouse models of neuroblastoma[15]. And cancer-causing versions of human papillomavirus include a gene that is not present in non-cancer-causing forms[16].

Many *de novo* genes remain uncharacterized, so the potential importance of the process to health and disease is unclear. "It will take some time before we fully understand to what extent it contributes to human health and to what extent it contributes to the origin of the human species," says Carvunis.

Although *de novo* genes remain enigmatic, their existence makes one thing clear: evolution can readily make something from nothing. "One of the beauties of working with *de novo* genes," says Casola, "is that it shows how dynamic genomes are." ■

---

**Adam Levy** *is a science journalist based in London.*

1. Baalsrud, H. T. *et al. Mol. Biol. Evol.* **35**, 593–606 (2018).
2. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
3. Begun, D. J., Lindfors, H. A., Thompson, M. E. & Holloway, A. K. *Genetics* **172**, 1675–1681 (2006).
4. Begun, D. J., Lindfors, H. A., Kern, A. D. & Jones, C. D. *Genetics* **176**, 1131–1137 (2007).
5. Albà, M. M. & Castresana, J. *Mol. Biol. Evol.* **22**, 598–606 (2005).
6. Carvunis, A.-R. *et al. Nature* **487**, 370–374 (2012).
7. Knopp, M. *et al. mBio* **10**, e00837-19 (2019).
8. Neme, R., Amador, C., Yildirim, B., McConnell, E. & Tautz, D. *Nature Eco. Evol.* **1**, 0127 (2017).
9. Zhang, L. *et al. Nature Ecol. Evol.* **3**, 679–690 (2019).
10. Toll-Riera, M. *et al. Mol. Biol. Evol.* **26**, 603–612 (2009).
11. Wu, D.-D., Irwin, D. M. & Zhang, Y.-P *PLoS Genet.* **7**, e1002379 (2011).
12. Casola, C. *Genome Biol. Evol.* **10**, 2906–2918 (2018).
13. Li, C.-Y. *et al. PLoS Comput. Biol.* **6**, e1000734 (2010).
14. Wang, D., Ma, J. Z. & Li, M. D. *Pharmacogenomics J.* **5**, 166–172 (2005).
15. Suenaga, Y. *et al. PLoS Genet.* **10**, e1003996 (2014).
16. Willemsen, A., Félez-Sánchez, M. & Bravo, I. G. *Genome Biol. Evol.* **11**, 1602–1617 (2019).

# COMMENT

ILLUSTRATION BY SEÑOR SALME

# Data — from objects to assets

How did data get so big? Through political, social and economic interests, shows **Sabina Leonelli**, in the fourth essay of a series on how the past 150 years have shaped science.

Data. The confusingly plural cornerstone of research. The grounding for a scientific understanding of the world. Lightning rods for the negotiation of political, social and economic interests.

Over the past 150 years, ideas have shifted drastically as to what counts as data, which data are reliable and who owns them. Once regarded as stable objects whose significance was determined by a handful of professional interpreters, data are now reusable goods. Their mettle depends on the extent to which they are mobilized across contexts and aggregated with others. Growing in volume, variety and value, data have come to drive the very process of discovery.

This explicit designation as assets has become possible only through a complex web of institutional, technological and economic developments. The history and consequences of how this web has been woven have repeatedly transformed research and its role in society.

150 nature

150 YEARS OF NATURE
Anniversary collection
go.nature.com/nature150

## COLLECTING COMMODITIES

Until the start of nineteenth century, efforts to collect facts and objects of study were spearheaded by visionary individuals, typically backed by wealthy patrons. Naturalists roamed the globe in search of biological specimens that were new to science. Court ▶

▶ astronomers devised tools to observe new parts of the cosmos. The large quantities of data accumulated were systematized and analysed through simple and powerful models (think Kepler's laws) and classification systems (such as that developed by botanist Carl Linnaeus). Thus was born the myth of the heroic theoretician, mining order from the chaos of observations. This individualistic view was tied to an understanding of data as fundamentally private — their scientific value residing in conceptual interpretation.

The nineteenth century marked a shift. Data, as we now recognize them, became institutionalized as social commodities. Their intellectual, financial and political worth arose from investments, requiring regulation and oversight. The botanical wonder cabinet that was Paris's natural-history museum was reorganized as a world-leading, publicly accessible repository of objects of potential scientific value. By the 1850s, the natural-history museums of Berlin, London and New York City followed suit.

The centralization of food markets spawned standardized approaches to the valuation and trade of organisms — such as the crop measures devised by the Chicago Board of Trade in Illinois. Cholera epidemics in Europe spurred large-scale collection of information on the spread and targets of disease. New methods of visualization and analysis emerged, such as physician John Snow's famous maps of how contaminated water spread cholera in central London.

National weather services started to build links between data collected regionally. The 1853 Brussels Convention on naval meteorology coordinated ships' logbooks into the first quasi-global data records for climate science. In Berlin, the first real bureau of standards, the Physikalische-Technische Reichsanstalt, was inaugurated in 1887 with physicist Hermann von Helmholtz as its founding director and a mandate to generate data needed for society as a whole. In the meantime, the US Army tasked the Library of the Surgeon-General's Office with collecting as many disease case reports as possible. Within 30 years, it had become the largest medical library in the world.

## NATIONAL TREASURES

By the turn of the twentieth century, the rise of nation states and the increasing demands of international trade drove initiatives to measure nature and society in a more systematic, objective way. National information infrastructures helped regions to share data, marking the start of a new informational globalism[1]. International entities, such as the League of Nations and the International Monetary Fund, yearned to globalize data collection and analysis for many purposes and across all scientific domains.

For example, the League of Nations Health Organization created the Permanent Commission on Biological Standardisation to monitor drug tests and biological assays from 1924. Well before the Second World War, there was increasing momentum to share information on employment, unemployment, wages and migration; from 1947, these data were amassed by the new International Statistical Commission. Such initiatives were fostered by an ever-expanding cadre of researchers, administrators, merchants and politicians.

All this fuelled the development of sophisticated approaches to quantification. Statistics emerged as a separate discipline — the main source of information for emerging insurance practices and public-health monitoring systems[2,3]. Techniques were developed to match the complexity of social exercises such as the census[4]. Population-level thinking gripped the life sciences, too — for good (genetics) and ill (eugenics). A new type of data collection focused on genetic mutants of a single model species[5,6], such as the fruit fly.

The two world wars severely disrupted data collection and sharing in the short term. But from the 1940s, the huge military investment in intelligence and information technologies kick-started the drive towards mechanized computing. The space race was perhaps the most notable cold-war contribution to globalized data systems and practices, particularly satellite technology. This produced the first global view of the planet and spurred the inauguration of the Intelsat system for worldwide civil-communications networks in the 1960s.

The World Meteorological Organization was founded in 1950 to oversee the international linkage of regional weather services, for instance in the Global Atmospheric Research Program. The International Geophysical Year of 1957–58 marked a step change in the commitment of Earth sciences to global data exchange, and was a diplomatic achievement in the middle of the cold war[7].

## GLOBAL GOODS

From the 1970s, almost every scientific field was building global, digitalized infrastructures for data sharing. The United Nations consolidated its global environmental monitoring system just as the World Health Organization systematized its efforts to map the spread of infectious diseases. The holy grail became the development of tools, such as computer models, that could crunch numbers at a previously unimaginable scale.

Increasingly, data were seen as sharable assets for repurposing, the value of which could change depending on their use. This view owed much to the cybernetics movement, with its emphasis on modularity and complexity[8]. Once again, the shifting role of data was also informed by the growth of international trade and the rising recognition of research as an engine of economic growth, military power and international relations.

Also in the 1970s, big science such as studies of particle collisions at Los Alamos National Laboratory in New Mexico and at CERN, Europe's particle-physics lab near Geneva, Switzerland, took centre stage. Here, the production and trade of data were no longer the responsibility of individual researchers. Rather, they were the output of large investment and collective efforts performed in centralized experimental facilities. Such centralization was unfeasible in many fields, for instance in environmental, biological and climate sciences, which work with observational rather than experimental data. Yet even those disciplines were focused on building networks for sharing

SSPL/GETTY

Microscopy slides used in the first detailed UK report of a link between lung cancer and asbestos.

PAUL WALTERS WORLDWIDE PHOTOGRAPHY/HERITAGE IMAGES/GETTY

**A Hollerith data machine at a steel works in Sheffield, UK, in 1963. The electromechanical device helped workers to tabulate statistics stored on punch cards.**

information so it could be fed into new computational tools.

Since the 1980s, portable computers, modelling and simulations have shaped data collection, manipulation and archiving. Climate scientists have developed ways to use legacy records to reconstruct a history of the atmosphere at the global level. This effort drove the pooling of international data, culminating in 1992 in the Global Climate Observing System.

In biology, the quest to map moved to the molecular level with big genetic sequencing projects, first in model organisms such as the nematode worm *Caenorhabditis elegans*, then through the Human Genome Project[9]. Sequencing databases were reimagined as playgrounds for discovery to facilitate immediate sharing, visualization and analysis online at a low cost, transforming the massive investment in genomic data production into useful knowledge.

## OPEN SEASON

As global data infrastructures and related institutions burgeoned, the resources needed to maintain them have mushroomed, and in ways that do not fit contemporary regimes of funding, credit and communication. For example, the curators of biological databases do essential work. But they do not routinely publish in top-ranking journals and might not be recognized or rewarded as high-level researchers. Similarly, keeping digital platforms robust and fit for purpose requires serious investment. The more data move around and are repurposed, the more vulnerable they are to unwarranted and even misleading forms of manipulation.

Over the past few decades, the Open Science movement has called for widespread data sharing as fundamental to better research. This has prompted several changes. One is the birth of journals devoted largely to the publication of data sets. Another is ambitious investment in data infrastructures, exemplified by the European Open Science Cloud. And the FAIR guidelines were crafted for how data should be labelled and managed to make them reusable[10]. There have also been calls to improve rewards for data stewards (such as technicians, archivists and curators), to raise their professional status from support workers to knowledge creators[11].

> *"The more data move around and are repurposed, the more vulnerable they are."*

These reforms are temporary solutions to a large-scale crisis of the contemporary research system, rooted in the inability to reconcile the diverse social and scientific aspects of data. The crisis recalls how the twentieth century reconfigured research data as political and economic assets. Their ownership can confer and signal power, and their release can constitute a security threat — as in the cold-war efforts to contain geological data that could have signalled nuclear testing. Now, new technologies are intersecting with emerging regimes of data ownership and trade. Starting from the 2000s, a handful of corporations has created — and wielded control over — new kinds of data left by billions of people as they meet, work, play, shop and interact online. (Think Amazon and Google.)

As algorithms become ever more opaque, the transparency and accountability of techniques and tools used to interpret data are declining. Whereas data curators remain the Cinderellas of academia, those who understand and control data management have climbed company ranks. And concerns are growing around data property rights, especially in the wake of misuses of personal data by the likes of Facebook and the UK company Cambridge Analytica.

Such tensions between data as public goods and private commodities have long shaped practices and technologies. Consider, for instance, the acrimonious debate over the ownership and dissemination of genomic data in the 1990s. On that occasion, free sharing won out through the establishment of the Bermuda Rules — an agreement among publicly funded researchers to deposit their sequences in public databases as soon as possible[12]. Wildly successful, this paved the way for open-data practices in other fields. Yet it also emphasized the financial advantages of owning genomic data[13,14] — a lesson swiftly learnt by companies that sequence and claim to interpret clients' genomes, which typically retain and use such data. Another example is the vast number of patents being filed for synthetic organisms by the chemical industries.

## VALUE ADDED

The use of big data as input for artificial-intelligence systems relies on the promise of global, comprehensive, easily available data riches. In principle, the marriage of powerful analytical tools with big biological data can support personalized medicine and precision agriculture. Similarly, social data hoovered up from Internet platforms and social-media services can inform evidence-based policy, business strategies and education. Yet history shows that moving research data around is not so simple. Underpinning technical questions around integration and use are thorny social, ethical and semantic issues.

How can different research cultures be encouraged to communicate effectively? What is the best way to collect, share and interpret data generated by the state, industry or social media? Which experts and stakeholders should have a say in data management and analysis? Who should have access to what, when and how? Addressing these issues requires effective administration and monitoring, and a long-term vision of the research domain at hand[15,16]. It also demands a repertoire of skills, methods and institutions geared to the study of specific research objects[17].

In summary, data generation, processing and analysis are unavoidably value-laden. The scientific legitimacy of these activities depends on the extent to which such values are held up for public scrutiny. Indeed, the best examples of data-intensive research to this day include strategies and methods to explicitly account for the choices made during data collection, storage, dissemination and analysis.

Model-organism databases such as PomBase (for the fission yeast *Schizosaccharomyces pombe*) and FlyBase (for *Drosophila*), for instance, clearly signal the provenance of what they store, including information about who created the data, for what purpose and under which experimental circumstances. Users can then assess the quality and significance of data[18]. Similarly, the Catalogue of Somatic Mutations in Cancer (COSMIC) captures the provenance of its holdings and the interpretive decisions taken by its curators while processing them. This helps clinicians to reassess the value of the information[19].

> ***"Extracting knowledge from data is not a neutral act."***

The more such assumptions and judgement are filtered by large digital infrastructures, the easier it becomes to hide or lose them, making it impossible for future generations to situate the data adequately. Data are cultural artefacts whose significance is clear only once their provenance — and subsequent processing — is known.

Technological development, particularly digitization, has revolutionized the production, methods, dissemination, aims, players and role of science. Just as important, however, are the broad shifts in the processes, rules and institutions that have determined who does what, under which conditions and why. Governance, in a word. Data emerge from this reading of history as relational objects, the very identity of which as sources of evidence — let alone their significance and interpretation — depends on the interests, goals and motives of the people involved, and their institutional and financial context. Extracting knowledge from data is not a neutral act.

Building robust records of the judgements baked into data systems, supplemented by explicit reflections on whom they represent, include or exclude will enhance the accountability of future uses of data. It also helps to bring questions of value to the heart of research, rather than pretending that they are external to the scientific process, as has arguably happened in bioethics[20]. This is a crucial step towards making big-data sciences into reliable allies for tackling the grave social and environmental challenges of the twenty-first century. ■

**Banks of servers at one of Google's US data centres.**

CONNIE ZHOU/GOOGLE/ZUMA PRESS

**Sabina Leonelli** *is professor of philosophy and history of science at the University of Exeter, UK.*
*e-mail: s.leonelli@exeter.ac.uk*

1. Hewson, M. in *Approaches to Global Governance Theory* (eds Hewson, M. & Sinclair, T. J.) Ch. 5 (State Univ. New York Press, 1999).
2. Desrosières, A. *La politique des grands nombres: Histoire de la raison statistique* (Editions La Découverte, 1993).
3. Porter, T. M. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton Univ. Press, 1995).
4. Von Oertzen, C. *Osiris* **32**, 129–150 (2017).
5. Kohler, R. E. *Lords of the Fly:* Drosophila *Geneticists and the Experimental Life* (Univ. Chicago Press, 1994).
6. Strasser, B. J. *Collecting Experiments: Making Big Data Biology* (Univ. Chicago Press, 2019).
7. Anorova, E., Baker, K. S. & Oreskes, N. *Hist. Stud. Nat. Sci.* **40**, 183–224 (2010).
8. Pickering, A. *The Cybernetic Brain: Sketches of Another Future* (Univ. Chicago Press, 2010).
9. Hilgartner, S. *Reordering Life: Knowledge and Control in the Genomics Revolution* (MIT Press, 2017).
10. Wilkinson, M. D. *et al. Sci. Data* **3**, 160018 (2016).
11. European Commission Directorate-General for Research and Innovation. *OSPP-REC: Open Science Policy Platform Recommendations* (European Union, 2017).
12. Maxson Jones, K., Ankeny, R. A. & Cook-Deegan, R. *J. Hist. Biol.* **51**, 693–805 (2018).
13. Parry, B. *Trading the Genome: Investigating the Commodification of Bio-Information* (Columbia Univ. Press, 2004).
14. Sunder Rajan, K. *Biocapital: The Constitution of Postgenomic Life* (Duke Univ. Press, 2006).
15. Edwards, P. N. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (MIT Press, 2010).
16. Daston, L. (ed) *Science in the Archives: Pasts, Presents, Futures* (Univ. Chicago Press, 2016).
17. Ankeny, R. A. & Leonelli, S. *Stud. Hist. Phil. Sci. A* **60**, 18–28 (2016).
18. Leonelli, S. *Data-Centric Biology: A Philosophical Study* (Univ. Chicago Press, 2016).
19. Forbes, S. A. *et al. Nucleic Acids Res.* **45**, D777–D783 (2017).
20. Leonelli, S. *Phil. Trans. R. Soc. A* **374**, 20160122 (2016).

ECONOMICS

# The clash of capitalisms

**Diane Coyle** compares three starry tomes on prosperity, justice and the environment.

**In Defense of Open Society**
GEORGE SOROS
*PublicAffairs (2019)*

**Capitalism, Alone: The Future of the System That Rules the World**
BRANKO MILANOVIC
*Belknap (2019)*

**Measuring What Counts: The Global Movement for Well-Being**
JOSEPH E. STIGLITZ, JEAN-PAUL FITOUSSI AND MARTINE DURAND.
*The New Press (2019)*

“In abnormal times the normal rules don’t apply and people obey them at their peril.” So writes ur-capitalist and philanthropist George Soros about the current “revolutionary moment” in *In Defense of Open Society*, his collection of essays diagnosing the sorry state of capitalism. Two other books with very different perspectives reach the same conclusion: that in the present tumult, the prevailing mode of economic organization is unsustainable.

In *Capitalism, Alone*, economist Branko Milanovic takes a less apocalyptic tone. Yet he shears away the woolliness of twentieth-century welfare capitalism. He reveals a future shaped by a clash of capitalisms — hyper-commoditized US versus hyper-politicized Chinese — and by polarization between elites and the rest. Meanwhile, in *Measuring What Counts*, Nobel laureate Joseph Stiglitz, with fellow economists Jean-Paul Fitoussi and Martine Durand, avoid big-picture pronouncements, focusing on the need for different metrics. But they respond to the same tensions, writing that we can no longer ignore “that so little of the fruits of growth over the past decades have gone to the bottom 90%”.

Given ongoing global disruption — from ‘populist’ governments to a burning Amazon rainforest — each book engages with politics as well as policies. Never a wholly technical subject, economics is returning to its roots as political economy. So, although none of the books proposes new solutions, each reflects that realignment. They could also be seen as a response to growing concerns about the health of capitalism after decades of short-term focus on financial gain alone.

## BETTER METRICS

Stiglitz and co-authors strongly advocate the use of economic indicators that dethrone gross domestic product (GDP, based broadly on the market value of all economic transactions). *Measuring What Counts* follows a 2009 report commissioned by the French government, translating its analysis into a policymakers’ manual for public consultation, and embedding “beyond GDP” metrics into budgets. The authors see statistics as central “to more democratic engagement” of societies, driving fundamental questions: “What do we value? What should we measure? What should we pay more attention to?”

As they remind us, anything unmeasured is invisible to policymakers. We are paying for the statistical invisibility of powerful trends, from human well-being to environmental degradation. Unequal incomes and health outcomes have not been measured regularly, so the declining life expectancy of US citizens without university degrees has become apparent only in the past few years. Many natural resources, worth some US$100 trillion to the global economy, have zero value in conventional statistical terms. No wonder they are wasted. All this is influenced by a 1997 paper by ecological economist Robert Costanza and others on ecosystem services and natural capital (R. Costanza *et al. Nature* **387**, 253–260; 1997).

Stiglitz, Fitoussi and Durand urge policymakers to use a dashboard of metrics emphasizing environmental, political and social sustainability. These are linked: environmental pressures trigger conflict and migration, for instance. The desire for comprehensive statistics encompassing global linkages lay behind the establishment of the 17 goals, 169 targets and 232 indicators of the United Nations Sustainable Development Goals. But these, the book rightly observes, are “too many to be meaningfully comprehended or to be a focus of policy”. By streamlining their dashboard, the authors make it more likely to be embedded in policy decisions.

Alternative metrics are a vigorous area of research, as attested by the World Bank’s measurement of ‘comprehensive wealth’ and New Zealand’s 2019 ‘well-being budget’. But this is not to say that the economics and policy worlds have, with the force of revelation, decided to ditch GDP. Many argue for improvement, not overthrow. One of the UN’s periodic revisions of the international System of National Accounts (SNA), which encompasses GDP, is just getting under way. Even so, today’s economic and political polarization, unavoidable evidence of environmental damage, and technological disruption of society have left the SNA fragile. It is more likely to be upended now than at any time since it was created in the late 1940s.

ANYTHING **UNMEASURED** IS **INVISIBLE** TO POLICYMAKERS.

With a better framework, things could be different, says Soros in *In Defense of Open Society*. His speeches and articles, written after the global meltdown of 2007–08, point to the liberal capitalist ideal: an ‘Open Society’, the dynamic, universalist, free society propounded by philosopher Karl Popper.

Like Stiglitz, Soros investigates the factors that contribute to instability. One, he argues, is the failure of orthodox economic theory, particularly the idea that the economic system tends to equilibrium. His 1987 book *The Alchemy of Finance* set out his alternative view as a successful investor, describing feedback loops in financial markets. But, as he notes, the economics establishment largely dismissed that as “the conceit of a man who has been successful in business and therefore fancies himself as a philosopher”.

It is hardly controversial, however. Soros posits that economic and social systems shape the views of people who participate in them, which affects their actions, in turn altering the economy. Start-ups might observe investors backing rapid growth at the expense of immediate profits, in a ‘winner-takes-all’ model. So they all adopt the same business system. This ‘reflexivity’ is equivalent to US sociologist Robert K. Merton’s principle of self-fulfilling outcomes. When you add uncertainty about the future, and lessons about human decision-making that are emerging from behavioural economics, the need for a subtler intellectual framework for economics is clear. The discipline is already shifting this way, with growing interest in psychology and social dynamics.

Another culprit threatens liberal capitalism, Soros argues: the technology industry, and the US–China tech ‘arms race’. He criticizes China’s authoritarian use of surveillance technology. Its developing ‘social credit’ system, for instance, aggregates personal data to rank citizens’ trustworthiness, affecting their access to certain rights. Soros is equally scathing about for-profit surveillance by the US behemoths of Silicon Valley. He likens it to extraction of natural resources; it’s just that “social media companies exploit the social environment”. Artificial intelligence represents mortal threats to liberalism because it puts power in the hands of a few, he argues.

## LIBERAL VS POLITICAL

In his overview of the recent history of capitalism and globalization, Milanovic likewise views current political and economic fragility in terms of a confrontation between China and the United States. *Capitalism, Alone* frames this as a conflict between models of capitalism: “liberal meritocratic” on the US side, “political” on China’s. Milanovic defines US liberalism as providing formally unrestricted opportunities for economic advancement. But given US inequities and lack of social mobility, ‘meritocratic’ seems debatable. Political capitalism, meanwhile, emerges from former communist countries, where the state’s needs take priority. The rule of law does not bind state action, and a technocratic bureaucracy is closely involved in

ILLUSTRATIONS BY ANTOINE DORE

economic management.

Milanovic contrasts both models with social democratic capitalism, shaped by economist John Maynard Keynes among others, including politicians rebuilding national economies after the Second World War. He also sketches an alternative "people's capitalism", echoed in today's European and US progressivism.

Milanovic was trained in the Marxist economics of the former Yugoslavia. This might explain why historical forces are so central to his analysis — and why he is not optimistic about the chances of alternative models. He lists policies that might favour progressivism, such as improving publicly funded education. In my view, these are wishful thinking without analysis of the political obstacles to sensible policies (sensible, at least, to those who prefer to avert revolution). Nor does the book explore other current varieties of capitalism, such as the more egalitarian, consensual systems of Scandinavia or Japan. It would have been salutary to know how these are responding to the forces of globalization, technological overreach, an ageing population and environmental stress.

These existential pressures explain why all three books conclude that 'business as usual' in thinking about how to run an economy cannot continue. It is right to measure what societies actually value, as Stiglitz and his co-authors (and I) argue. It is also right, as Soros asserts, that the intellectual framework of economics must adapt to a world ever more removed from a focus on individual choices. This trend is under way in economic research, but a radical rethink is unlikely there: the incentives of academia encourage conservatism and incremental progress.

Better metrics and theories will not be enough to create a sustainable economic and social model. Or, they could — but only if they convince policymakers and the public to act differently. The future of capitalism is out of the hands of those who spend their time thinking about it. ■

---

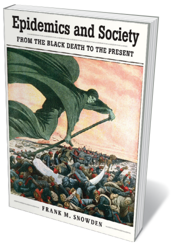
**Diane Coyle** *is Bennett Professor of Public Policy at the University of Cambridge, UK.*
*e-mail: dc700@cam.ac.uk*

---

EPIDEMIOLOGY

# How pandemics shape social evolution

**Laura Spinney** weighs up Frank Snowden's sweeping history of the impact of infectious diseases on society.



**Epidemics and Society: From the Black Death to the Present**
FRANK M. SNOWDEN
*Yale University Press (2019)*

When will we learn never to declare the end of anything? Only 50 years ago, two prominent US universities closed their infectious-disease departments, sure that the problem they studied had been solved. Now, cases of measles and mumps are on the rise again in Europe and the United States, new infectious diseases are emerging at an unprecedented rate, and the threat of the next pandemic keeps philanthropist Bill Gates awake at night.

So it's a shame that to make this point, *Epidemics and Society*, Frank Snowden's wide-ranging study on this rolling human reality, repeats the urban myth that in 1969, US surgeon-general William Stewart said, "It is time to close the book on infectious diseases, and declare the war against pestilence won." Even though Stewart never said this, it's clear that there was a pervasive, dangerously complacent attitude in the late 1960s. International public-health authorities were predicting that pathogenic organisms, including the parasite that causes malaria, would be eliminated by the end of the twentieth century. Snowden's broader thesis is that infectious diseases have shaped social evolution no less powerfully than have wars, revolutions and economic crises.

SCAPEGOATING AND **VIOLENCE** HAVE ACCOMPANIED **EPIDEMICS.**

It's not a new message, but it bears repeating. Snowden, a historian at Yale University in New Haven, Connecticut, has assembled a vast amount of evidence, some the fruit of his own research. His global history spans more than a millennium of outbreaks, covering diseases from bubonic plague to smallpox, malaria, the respiratory illness SARS, Ebola and beyond. He rehashes the long history of scapegoating, violence, mass hysteria and religiosity that have accompanied epidemics, but only to speculate on their longer-term social, political and cultural consequences.

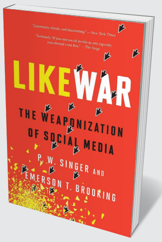
When cholera struck Paris in 1832 — in an epidemic that eventually killed nearly 19,000 Parisians — a conspiracy theory spread that the unpopular government under King Louis Philippe was poisoning wells with arsenic. The police and army were barely able to contain the violence that ensued. The institutional memory of those events fuelled dread of the "dangerous classes": poor people. That, Snowden argues, might help to explain why the two most egregious examples of class-based repression in the nineteenth century also took place in the French capital. These were the violent crushing of the 1848 revolution and the bloody destruction of the Paris Commune, the revolutionary government that briefly ruled the city 23 years later.

The synergy between wars and epidemics in shaping history has long been recognized. Napoleon Bonaparte's nineteenth-century

---

NEW IN PAPERBACK

*Highlights of this season's releases.*



**LikeWar**
*P. W. Singer & Emerson T. Brooking* HOUGHTON MIFFLIN HARCOURT *(2019)*
As reports on politics and war flood social media, the medium itself is becoming weaponized: virality is valued over veracity. If you're online, you inadvertently become part of the war. Warning that 'you are what you share', defence specialists P. W. Singer and Emerson Brooking explore the real-world and online geopolitical impacts of this conflict, and how to prepare ourselves for the next unprecedented threat.

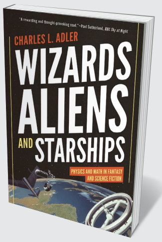
imperial expansion westwards across the Atlantic Ocean was halted by yellow fever, which his army encountered in France's Caribbean colony of Saint-Domingue (now Haiti). His eastern ambitions were thwarted by dysentery and typhus. (The typhus epidemic that ravaged the Grande Armée during its retreat from Moscow might have prompted an unparalleled die-off by some measurements, as Snowden claims. But it was surely not in terms of "deaths per capita".)

## SECURITY THREAT

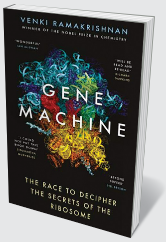
An odd omission from the book is the 1918 'Spanish' influenza pandemic, which overlapped with the First World War and is estimated to have killed between 50 million and 100 million people. Snowden might have felt that it garnered enough attention around its centenary. But a future flu pandemic is currently ranked among the leading threats to global security, and there has been surprisingly little research on the long-term consequences of the 1918 catastrophe. Furthermore, it might have been interesting to explore the possible links between that pandemic and the ongoing epidemic of AIDS in South Africa, which the book does cover.

There is evidence that white scapegoating of black South Africans in 1918 precipitated the first legislative steps towards apartheid. As Snowden discusses, by restricting the land available to people of colour, apartheid accelerated the growth of a migrant labour system that divided black families. It also encouraged new forms of social and sexual behaviour. Both developments, in turn, hastened the spread of AIDS once it arrived. Young men growing up away from their families, for instance, often developed standards of masculinity that promoted sexual conquest and violence; South Africa now has one of the highest rates of rape in the world (K. Naidoo *S. Afr. Med. J.* **103**, 210–211; 2013). The crowning tragedy of these depressing events was President Thabo Mbeki's endorsement, from 1999, of a US AIDS denialist's theory that the disease is not caused by the HIV virus. That resulted in the avoidable deaths of an estimated half a million South Africans.

In the twenty-first century, we seem to be repeating many of the mistakes that triggered or exacerbated epidemics in the past. That is presumably why Snowden refers ominously to this century's first major epidemics, of



**Wizards, Aliens and Starships**
*Charles L. Adler* Princeton Univ. Press *(2019)*
Physicist Charles Adler tracks his own field and maths through science-fiction tropes to separate the plausible from the impossible. The shape-shifting transfiguration spells in J. K. Rowling's Harry Potter series, for instance, pose problems related to conservation of mass.



**Gene Machine**
*Venki Ramakrishnan* Oneworld *(2019)*
This scientific memoir by UK Nobel laureate and Royal Society president Venki Ramakrishnan is invitingly witty. He gives a frank account of the race to demystify the ribosome (the cell's protein factory), and the highs and frustrations of scientific success (see G. Ferry *Nature* **561**, 32; 2018).

SARS and Ebola, as “dress rehearsals”. Although many people espouse health care for all, our globalized economic system militates against it — because profits are rarely invested where they were extracted — and we still seem to think that borders will keep disease out, even though they never have. Since Snowden completed his book, the administration of US President Donald Trump has announced that an immigrant’s chances of getting permanent residence will now be linked to the burden they put on the public purse — including health-care costs. That makes it more likely that recent arrivals will avoid doctors, and infectious diseases will go undetected.

The starkest reminder that the battle is not won, however, is that only one infectious disease has been eradicated globally: smallpox. Others that those optimists of the 1960s thought would have vanished by now have been hard to dislodge — and could easily flare up again. The strife-ridden Democratic Republic of the Congo is harbouring more than Ebola. There is also a measles outbreak, and a circulating strain of polio that mutated from the live, weakened one in the oral vaccine. There have been successful local disease eradications, but they often came at a price. A sustained campaign of DDT application helped eliminate malaria from the Italian island of Sardinia by 1952, for example, but in 2001, the pesticide was banned globally under the Stockholm Convention, after it was found to be dangerous to wildlife and the environment.

For Snowden, the lesson from more than 50 years of such experiments — successes and failures — is that eradication is most likely to work when doctors, politicians, drugmakers, the media and citizens work together. *Salus populi suprema lex*, he reminds us: public health must be the highest law. He has preached that message to generations of Yale undergraduates, and repeats it in this book. The risk is only that he is preaching to the converted. ■

---

**Laura Spinney** *is a science writer based in Paris. Her most recent book is* Pale Rider: The Spanish Flu of 1918 and How it Changed the World.
*e-mail: lfspinney@gmail.com*

HISTORY OF TECHNOLOGY

# Shadowed light

**David E. Nye** examines volumes on the complex history of lighting technologies and a great inventor.

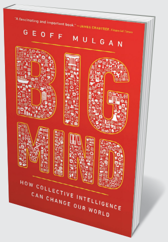
**American Lucifers: The Dark History of Artificial Light, 1750-1865**
JEREMY ZALLEN
*Univ. North Carolina Press (2019)*

**Edison**
EDMUND MORRIS
*Random House (2019)*

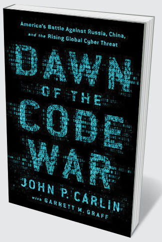
Beyond steam engines and power looms, the Industrial Revolution spawned innovation in artificial lighting — for city streets, lighthouses, railway carriages and mills. Incremental improvements, from whale oil to gas, kerosene and electric carbon-arc technology, culminated in 1879 in a practical incandescent light bulb created by prolific US inventor Thomas Edison. These advances were no straightforward march of progress, as two books — Jeremy Zallen’s *American Lucifers*, and *Edison* by the late Edmund Morris — reveal.

*American Lucifers* begins a century before Edison’s birth, in 1750. Zallen, a historian, explores the human costs of artificial lighting from then until 1890, concentrating on people producing and using fuel in the United States, and their links to South America, Britain and the Caribbean. Viewing energy in terms of class, he examines the fate of whalers, enslaved people distilling pine resin to make turpentine, children in match factories, petroleum refiners and miners of coal and copper.

The international networks he examines are convoluted. Jewish candle-makers in



**Big Mind**
*Geoff Mulgan* PRINCETON UNIV. PRESS *(2019)*
Innovation specialist Geoff Mulgan’s timely work draws on philosophy and computer science to explore collective intelligence: how combining human and technical abilities could help to tackle everyday problems, along with large-scale challenges in public health and climate change.



**Dawn of The Code War**
*John P. Carlin & Garrett M. Graff* PUBLICAFFAIRS *(2019)*
This cautionary insider story by security strategist John Carlin and journalist Garrett Graff examines targeting of US interests in cyberspace. From election hacking to terrorist recruitment, they provide legal insight into the risky situation facing the United States online.

Rhode Island, he argues, were culpable for the exploitation of workers who used their candles in Caribbean sugar plantations, New England cotton mills and Pennsylvania coal mines. Later, coal from those mines fed steam engines that powered electric lights. The candles themselves were made using either whale oil or tallow from livestock — cattle and hogs that "until the 1930s were mostly raised by farmers using kerosene lamps". Zallen's case studies also range over Argentinian slaughterhouses, phosphorus-match factories in Liverpool, UK, and a Montana copper mine.

The benefits of artificial light get short shrift. The "whole electric edifice", Zallen argues, was predicated on "a century of steadily industrializing slave labor" along with the exploitation of women and children who, although not enslaved, worked at best for starvation wages. Yet such assertions oversimplify. In the nineteenth century, real wages in the United States increased. Factories demanded literacy and numeracy, which is one reason children were compelled to attend school. Nor is women's history during this era one purely of exploitation and "rigid expectations", as the spread of women's colleges and the growing suffrage movement reveal.

In arguing that electrification was dangerous for miners, Zallen fails to acknowledge that, after around 1890, electric light proved safer in mines than candles. Electric-fan ventilation removed explosive gases, and alarms and telephones improved safety. But such technologies did not necessarily raise wages, and they did lead to redundancies.

Zallen is right to stress the human costs of resource extraction and to see technologies as ideological by implication and often oppressive in use. Energy historians need to integrate labour with invention and entrepreneurship, and *American Lucifers* contributes valuable perspectives. However, this well-written work comes close to reducing a complex transition between energy regimes into a simpler story of class struggle.

The packed narrative shows that the mid-nineteenth-century world into which Edison was born was not simply a prelude to electrification. But Zallen mistakenly dismisses the inventor as an "electric booster", when he was actually well acquainted with the energy world of candles, coal and class conflict. Morris's view of Edison is more nuanced. Known for his trilogy on US president Theodore Roosevelt, Morris follows an unusual chronology, proceeding in reverse from Edison's death in 1931. Thus, Edison's final three decades, when important inventions tapered off, precede his spectacular triumphs.

Born in 1847 in rural Ohio, and moving to Michigan as a child, Edison had little schooling before he became a telegrapher at 15. He learnt how the telegraphy system worked, and became an inventor, ultimately based in New Jersey. At 30, he had improved telegraphy, the telephone and the microphone, and startled the world with the phonograph. By 1882, he and his collaborators had developed the electric bulb and the still-familiar electricity-distribution system of dynamos, wiring, fuses, sockets and wall switches.

Electric lighting outshone gas first in city centres, stock exchanges, railway stations and prominent buildings; by 1910, it dominated street lighting. By 1940, it was on its way to today's vast, energy-guzzling infrastructure. Morris, an engaging writer with an eye for details, explains the inventions clearly, including lesser-known technologies such as the 1877 translating embosser, which sped up telegraphic transmission. Perhaps Edison's greatest invention was the first US industrial research lab, in West Orange, New Jersey. From 1886, this developed motion-picture equipment and a film studio; improved batteries, iron mining, cement production and phonograph recording; and found that a weed, goldenrod (*Solidago leavenworthii*), could be a domestic source of rubber.

ELECTRIFICATION HAS **PROVED A BOON** BUT SPURRED RESOURCE EXTRACTION AND **GLOBAL WARMING.**

Edison — twice married and a conflicted father to six children — was more devoted to the lab than to domestic life. Invention was a collective process, as the thousands of collaborative notebooks held at West Orange attest. His relationships with pioneering electrical engineer William Hammer, his chief experimental assistant Charles Batchelor, and others were essential to Edison's success, but in later life he became more of a loner.

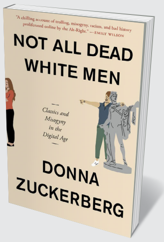
Like many inventors, he proved poor at business. Like many a self-made man, he had little sympathy for workers in the difficult 1890s. And like many founding entrepreneurs, he resisted delegation. At his death, his famous laboratory was left moribund and leaderless. Edison had long since lost control or even partial ownership of his electric-light interests, which financiers submerged in the General Electric company. He let major inventions, such as the phonograph and motion pictures, languish, but poured resources into quixotic projects — notably, a New Jersey iron mine, to the despair of family and associates. By telling Edison's story in reverse, Morris downplays these accumulated failings, in a riches-to-rags narrative that explores how individual creativity emerges.

Morris's treatment is detailed but flawed. Edison's relationship with the US car magnate Henry Ford was more important than Morris credits: both grew up in the hinterlands of Detroit, Michigan, and they held similar convictions. Edison's racism is only gingerly mentioned. Some fascinating letters to Edison, including several from the escape artist Harry Houdini, might have been included. But, faced with five million pages of documents, Morris had to omit much. The structure also inhibits retrospective summaries.

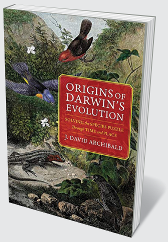
Electrification has proved a boon — but spurred resource extraction, light pollution and global warming. As we face another great transition, from fossil fuels to alternative energies, Zallen's narrative is timely — echoing in the high human and environmental costs of dramas playing out in Nigerian oilfields and the smog of Indian cities. Meanwhile, Morris's portrait of an energy revolution that was eagerly embraced, yet took more than half a century, leaves us hoping that rendering Edison's system sustainable will not take as long. ■

---

**David E. Nye** *is professor of US history at the University of Southern Denmark in Odense. His publications include* Electrifying America, When the Lights Went Out *and* American Illuminations.
*e-mail: nye@sdu.dk*



**Not All Dead White Men**
*Donna Zuckerberg* Harvard Univ. Press *(2019)*
With the proliferation of anti-feminist rhetoric online, the extreme right is using ancient philosophy to boost its credibility. As Stoic ethics moves from lecture halls to Reddit, classicist Donna Zuckerberg exposes this misappropriation, meant to enforce the concept of male superiority.



**Origins of Darwin's Evolution**
*J. David Archibald* Columbia Univ. Press *(2019)*
Biologist David Archibald examines an unsung hero of Charles Darwin's evolutionary theory: historical biogeography, the natural history of species in time and place. Archibald invites us to enrich our understanding through Darwin's ideas on species evolution in different regions.

MARINE SCIENCES

# The ocean re-imagined

Two books offer urgent takes on the state of the seas. **Boris Worm** lauds both.

**Neptune's Laboratory: Fantasy, Fear, and Science at Sea**
ANTONY ADLER
*Harvard University Press (2019)*

**Reef Life: An underwater memoir**
CALLUM ROBERTS
*Profile (2019)*

On a calm day, the ocean can resemble a vast mirror. Peering over the side of a boat, we might see ourselves reflected; what lies beneath is hidden.

In *Neptune's Laboratory*, environmental historian Antony Adler takes this observation as a leitmotif. The ocean, he writes, is "an ideal screen for human projections of fear and hope". In his entertaining, readable history of marine science, the author shows how humanity's fundamental ignorance about the sea has often fed fantastical ideas of it as saviour, battlefield, playground, storehouse, angry beast or hapless victim. Throughout, he reminds us, we have struggled to see Earth's oceanic reaches for what they truly are: the face of our changing planet.

That recognition of a rapidly, irreversibly altering ocean permeates every page of Callum Roberts's scientific memoir, *Reef Life*. He takes a deep dive into his own four-decade career as a marine ecologist, chronicling the splendour, complexity and vulnerability of coral reefs. Both books left me with a sense of urgency about the ocean's perilous state, but also with renewed hope that we have reached a turning point in our collective relationship with it.

Through his eventful tale, Adler recounts how scientific inquiry into the ocean began in earnest less than 200 years ago, and how the findings of myriad individuals gradually coalesced into an interdisciplinary field: oceanography. Adler discusses many colourful personalities. For instance, the "Prince of Ocean Science" — Albert I of Monaco — funded the early expansion of oceanography in Europe in the late nineteenth and early twentieth centuries. And French under-sea explorer Jacques Cousteau popularized the ocean through film and television.

Two opposing trends feature in *Neptune's Laboratory*. On the one hand, marine scientists such as Albert I have often worked hard to forge international collaborations, recognizing the ocean as a common good without clear boundaries, too large for any one nation to grasp. Others have tended towards nationalism and parochialism, seeking to carve up the ocean for political, economic or military gain. One such was US mechanical engineer Carroll Livingston Riker. In 1912, he unsuccessfully lobbied the US Congress to spend US$190 million on a 320-kilometre jetty, intended to redirect the Labrador Current and Gulf Stream, warming the Arctic to produce ice-free harbours.

## A RICHER CONTEXT

Personally, I found Adler's study enlightening. I have rarely seen the history of oceanography taught in any comprehensive way in marine-science classes, either ones I have taken or those I have given. Reading the book, I came to see my own journey as a marine scientist in a much richer context, tightly interwoven with the personalities, philosophies and storylines that dominate the field's history.

Of particular resonance, for example, is the transformation in our understanding of oceanic limits over the past 150 years. All too recently seen as an inexhaustible, boundless resource — bolted to the hubristic sense that the 'seven seas' could be controlled and conquered — Earth's ocean ecosystems are now being recognized as fragile and in decline. More than 90% of fish stocks are fully or overexploited, and climate change is compromising the ocean's oxygen supply and productivity. Adler quotes oceanographer Sylvia Earle: "nothing else will matter if we fail to protect the ocean. Our fate and the ocean's are one."

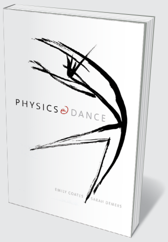
That final narrative and call to action largely dominates Roberts's *Reef Life*. This is a deeply personal journey of a marine scientist and conservationist whose working life takes place in the oceans of the Anthropocene — the geological epoch proposed to mark humanity's dominant impact on planet Earth. Throughout his travels, from the Red Sea and the Gulf to remote atolls in the Pacific, Roberts witnesses the slow disintegration of coral reefs from the combined impacts of land-based pollution, habitat destruction, overfishing and ocean warming.

Casting coral reefs as a canary in the coal mine, Roberts warns of the transformative effects of climate change and other human stressors on oceans. Part odyssey, part 'Reef Ecology 101', Robert's witty and vivid descriptions of the underwater world are meshed with the most up-to-date findings, which suggest that "reefs cannot be climate-proofed and they can't hide from climate change". Without drastic emissions reductions, we are indeed facing a world without coral.

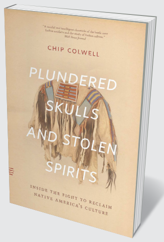
## BATTERED REEFS

Some of his imagery is haunting. He compares the battered and bleached reefs he has witnessed through his career to ailing parents gradually robbed "of mobility, independence and even dignity". I can relate to Roberts's painful transition from bright-eyed student fascinated by marine biodiversity to chronicler of decline and advocate for a saner relationship with the sea. Many in our field have watched what we love disappear before our eyes, but few have recounted it in such a public way.

Both books reveal geographical limitations. *Neptune's Laboratory* is almost exclusively set in Britain, France and the United States. Yet Adler points out that Russia's marine-science history is understudied and underappreciated, partly because much of it happened under a veil of secrecy



**Physics and Dance**
*Emily Coates & Sarah Demers* YALE UNIV. PRESS *(2019)*
New York City Ballet dancer Emily Coates and CERN physicist Sarah Demers interweave science and choreographic research in this unique study. That fusion, they show, can enrich understanding of both fields. An insightful *pas de deux* between physics and ballet.



**Plundered Skulls and Stolen Spirits**
*Chip Colwell* UNIV. CHICAGO PRESS *(2019)*
The fight to reclaim Native American culture goes on. Here, curator Chip Colwell sensitively explores repatriation of human remains held in museum collections, including the scalp of a Native American murdered by the US Army in Colorado's 1864 Sand Creek massacre.

during the cold war. Likewise, members of non-European maritime cultures including Polynesians and Inuit people have long built a deep understanding of the ocean, and framed their relationship to it in different, and possibly more holistic ways, but these are not mentioned here.

And in *Reef Life*, Roberts's global travels are viewed very much through his own cultural lens. There is little perspective on the lives and views of the local people who are most affected by the changes he chronicles.

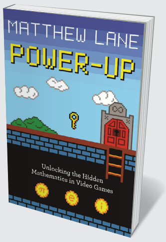
Yet these books both offer a valuable reminder that we are at a crossroads in our collective relationship to world oceans. Humanity now has an unprecedented awareness of what the ocean does for us, such as regulating Earth's climate though heat absorption and carbon sequestration. At the same time, we increasingly understand how human impact is disrupting ocean systems and threatening the abundance of ocean life. With that understanding comes the opportunity to allow marine ecosystems to recover, as detailed, for example, in the 14th United Nations Sustainable Development Goal, 'Life Below Water'.

NOTHING ELSE WILL
**MATTER**
IF WE FAIL TO
**PROTECT**
THE
**OCEAN.**

I wondered, after reading *Neptune's Laboratory* and *Reef Life*, whether we are on the cusp of re-imagining the ocean once more when we dive through that mirroring surface to comprehend its depths. As marine scientists Jane Lubchenco and Steven Gaines (*Science* **364**, 911–912; 2019) have put it: "The ocean is not too big to fail, nor is it too big to fix. It is too big to ignore." ■

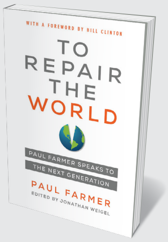
---

**Boris Worm** *is a marine ecologist and Killam Research Professor at Dalhousie University in Halifax, Canada.*
*e-mail: boris.worm@dal.ca*



### Power Up

*Matthew Lane* Princeton Univ. Press *(2019)*
Passionately nerdy mathematician Matthew Lane reveals how video games could be useful teaching aids. He explores the not-so-hidden maths in the classics — for example, using *Mario Kart* turtle shells to elucidate the study of pursuit and evasion trajectories. Enlightening.



### To Repair The World

*Paul Farmer* Univ. California Press *(2019)*
This heartfelt manifesto calls for the next generation to tackle challenges ranging from climate change and health-care access to essential human rights. With humour and passion, medical anthropologist Paul Farmer advocates a cure for society and the planet.

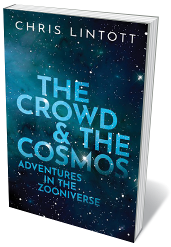
ASTRONOMY

# Putting the 'I' in science

Chris Lintott's chronicle of the booming citizen-science project Zooniverse is inspirational, finds **Michael West**.

Citizen science is booming. Today, anyone with a computer or a smartphone can participate in research in astronomy, oceanography, medicine, zoology and beyond. With such studies no longer the exclusive realm of an elite few, communities of amateur and professional scientists have joined together to democratize the discipline, harnessing mutual enthusiasm and collective wisdom to gather and analyse data.

As a research tool, crowdsourcing is nothing new. Charles Darwin maintained a voluminous correspondence with fellow naturalists and lay enthusiasts in the Victorian era. For more than a century, the US-based National Audubon Society has relied on an army of volunteers to count birds across North America each December. And since 1911, the American Association of Variable Star Observers in Cambridge, Massachusetts, has enlisted a network of predominantly amateur astronomers to collect nearly 40 million observations of stars that have fluctuating brightness. That endeavour has provided valuable insights into stellar lifecycles and distances to galaxies. SETI@home, launched in 1999, meanwhile uses the idle time on millions of home computers to search for radio signals from extraterrestrial civilizations.

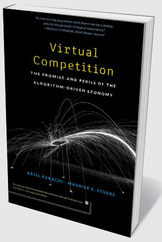


**The Crowd and the Cosmos: Adventures in the Zooniverse**
CHRIS LINTOTT
*Oxford University Press (2019)*

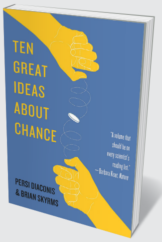
In recent years we've seen an explosion in new opportunities, in fields such as cetology, linguistics and space archaeology. The SciStarter website (https://scistarter.org), for example, currently aggregates thousands of citizen-science projects and events from around the world.

One researcher who has been at the forefront of the phenomenon for more than a decade is Chris Lintott, professor of astrophysics at the University of Oxford, UK, and a presenter of *The Sky at Night*, the BBC's monthly astronomy television show. In his new book, *The Crowd and the Cosmos*, Lintott tells the story of the most ambitious, successful citizen-scientist initiative so far: Zooniverse, which boasts 1.6 million registered users. Through its platform, people can, in effect, become research assistants to scientists working on projects in a profusion of fields. What these have in common are large quantities of data and a need for human eyes, ears and brains to help make sense of them.

Zooniverse emerged from the success of Galaxy Zoo, Lintott's first citizen-science venture. In 2007, faced with the daunting task of classifying millions of galaxies imaged by the Sloan Digital Sky Survey telescope in New Mexico, Lintott and colleagues solicited help through a brief slot on BBC Radio's morning



**Virtual Competition**
*Ariel Ezrachi & Maurice E. Stucke* HARVARD UNIV. PRESS *(2019)*
From price-comparison algorithms to phone operating systems, technology has altered competitive commerce. Lawyers Ariel Ezrachi and Maurice E. Stucke question the democratic consequences of this dual-edged power.



**Ten Great Ideas About Chance**
*Persi Diaconis & Brian Skyrms* PRINCETON UNIV. PRESS *(2019)*
Philosopher Brian Skyrms and mathematician Persi Diaconis weave the foundations of probability with economics and history in this engrossing discourse. A must-read for anyone interested in the dissection of probability. **Mary Craig**

current-affairs programme. The response was beyond anyone's wildest dreams. Within a few days, volunteers were classifying 70,000 galaxies every hour on their own computers.

One of Lintott's key messages is that citizen science is much more than free labour. Many such projects exploit the human brain's ability to recognize patterns, or to spot unusual features in data that even the most sophisticated computer algorithms can miss. Collaborations between professional and amateur researchers also increase public understanding of science, and have produced a growing list of publications in peer-reviewed journals. The first Galaxy Zoo paper, released in 2008, has been cited in more than 500 other astronomy papers (C. J. Lintott *et al.* Preprint at https://arxiv.org/abs/0804.4483; 2008). Reef Life Survey, a citizen-science project that engages recreational divers around the world to monitor biodiversity in coral reefs, has produced nearly 60 peer-reviewed papers so far, including 5 in this journal.

Citizen scientists have also made serendipitous discoveries on their own. In 2007, for example, Dutch school teacher Hanny van Arkel stumbled upon a mysterious green blob in an image she was examining for Galaxy Zoo. This unusual object, which became known as Hanny's Voorwerp (Dutch for 'Hanny's thingy'), is now thought to be a giant cloud of gas illuminated by a powerful blast from a supermassive black hole in the neighbouring galaxy IC 2497.

Zooniverse, as Lintott shows, hugely expands the field of investigation. Penguin Watch, for example, invites volunteers to monitor the rise and fall of Antarctic penguin populations by counting birds photographed by a network of automated cameras. Snapshot Serengeti uses a similar approach to study animal ecosystems revealed by millions of photographs taken with motion-sensitive cameras throughout Tanzania's Serengeti National Park. The Space Warps project invites armchair astronomers to search for rare but spectacular gravitational lenses, created when gravity distorts images of faraway galaxies. These act like enormous funhouse mirrors to produce optical illusions on the grandest scale.

Lintott is not the first to write about this topic. Caren Cooper's 2016 *Citizen Science*, for example, was illuminating. However, it is hard to imagine anyone more qualified than Lintott — a veteran of the citizen-science trenches — to give an insightful perspective. And he does so both accessibly and engagingly. There is a flavour of Bill Bryson's breezy erudition in *A Short History of Nearly Everything* (2003), although the book does ramble in places. Overall, however, Lintott deftly interweaves personal experience and more philosophical ruminations on public participation in science.

WITHIN A FEW DAYS, VOLUNTEERS WERE CLASSIFYING 70,000 GALAXIES EVERY HOUR ON THEIR OWN COMPUTERS.

What of the future of citizen science? Astronomy, once photon-starved, will soon be awash in 15–30 terabytes of new data nightly from the Large Synoptic Survey Telescope in northern Chile, triggering myriad follow-up observations. Other fields face similar challenges coping with an ever-faster flow of data. Genomic researchers are both blessed and burdened by a deluge of data emanating from hugely accelerated sequencing. And remote-sensing observations by a growing armada of satellites, such as the joint European–Japanese EarthCARE mission scheduled for launch in 2021, will map and measure our planet's surface as never before over the next decade. Although increasingly powerful computers and artificial intelligence can help to analyse the data tsunamis, they won't make citizen scientists obsolete any time soon.

Scrutinizing Earth's surface is one thing; having an impact on the future of the planet and its people is another. Can citizen science change the world? Maybe. From monitoring flower production by plants as a gauge of climate change to analysing brain scans in the quest to find the cause of Alzheimer's disease, lay researchers are actively improving lives globally. Just days after Hurricane Dorian devastated the Bahamas in late August, a new Zooniverse project was already helping rescue efforts as volunteers identified damage visible in satellite images.

Moreover, as Lintott reminds us, this great public venture is helping to foster a more scientifically literate society, and empowering a new generation of scientists. Not bad for a free app that you can download to your phone. ■

---

**Michael West** *is Deputy Director for Science at Lowell Observatory in Flagstaff, Arizona, USA. He is also Secretary of the International Astronomical Union's commission on Communicating Astronomy with the Public.*
*e-mail: mwest@lowell.edu*

---

CHEMISTRY

# From bomb to Moon

## **Angela N. H. Creager** is inspired by the life of the Nobel laureate who discovered deuterium.

After witnessing the 1945 Trinity atomic-bomb test, the theoretical physicist J. Robert Oppenheimer recalled Hindu scripture: "Now I am become Death, the destroyer of worlds." Although this is often interpreted as admitting moral culpability on the part of the Manhattan Project's scientific director, Oppenheimer remained a central player in the nuclear-weapons establishment until he lost his security clearance in the mid-1950s.

Harold Urey also worked for the Manhattan Project. But by contrast, the Nobel-prizewinning chemist distanced himself from nuclear weapons development after the war. His search for science beyond defence work prompted a shift into studying the origins of life and lunar geology. Now, the absorbing biography *The Life and Science of Harold C. Urey* by science historian Matthew Shindell uses the researcher's life to show how a conscientious chemist navigated the cold war.

Shindell argues that Urey's pious upbringing underpinned his convictions about the dangers of a nuclear arms race, and his commitment to research integrity. Urey grew up a minister's son in a poor Indiana farming family belonging to a plain-living Protestant sect, the Church of the Brethren. Progressing through increasingly diverse educational environments, culminating in a PhD at the University of California, Berkeley, Urey became self-conscious about the zealousness of his family's faith. He also found the path to a cosmopolitan, middle-class life.

In the 1920s, Urey was among a small group of chemists who collaborated closely with physicists. Working at Niels Bohr's

Institute for Theoretical Physics at the University of Copenhagen, he kept abreast of developments in quantum mechanics. There, and on travels in Germany, he met the likes of Werner Heisenberg, Wolfgang Pauli and Albert Einstein. But Urey decided he lacked the mathematical skills to make theoretical advances in quantum chemistry. Moving back to the United States, he started both a family and an academic career.

At Johns Hopkins University in Baltimore, Maryland, and later at Columbia University in New York City, Urey taught quantum mechanics to chemists, while setting out on the trail that led him to deuterium. In 1931, he discovered this isotope of hydrogen. Predicted on the basis of work by Bohr, Frederick Soddy, and J. J. Thomson, its existence had been doubted by many chemists and physicists. Urey's identification won him the Nobel three years later. By this time, he had also co-authored one of the first texts in English on quantum mechanics as applied to molecular systems, the 1930 *Atoms, Quanta and Molecules*.

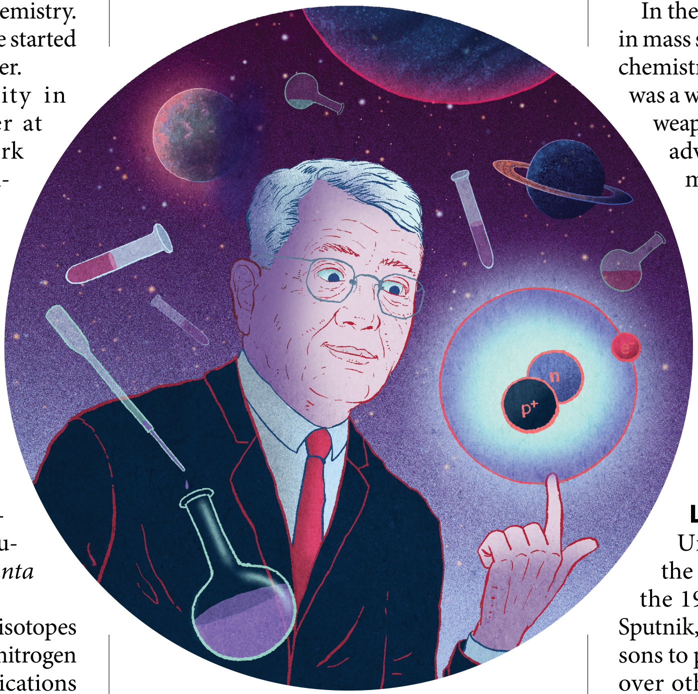
Urey's continuing work on stable isotopes of other chemical elements, such as nitrogen and oxygen, led to important applications in biochemistry and geochemistry, including the pioneering use of isotopic labels to study metabolic pathways. Living in New York also led Urey to political liberalism. He became aware of the anti-Semitism affecting Jewish scientists, and the lack of opportunities for women scientists. A generous mentor, he shared his Nobel prize money with two collaborators, and split a grant he had been awarded with the young Isidor Rabi (who later discovered nuclear magnetic resonance).

## MANHATTAN TRANSFER

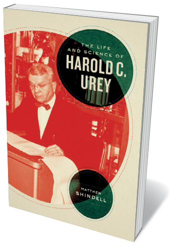
The Second World War changed Urey's life, as it did those of most physical scientists and researchers in many countries. His expertise in isotopes made him valuable to the Manhattan Project. Here, he eventually headed a massive team of scientists and engineers working on the separation of uranium isotopes using gaseous diffusion methods. However, he was ill-suited to the pressure of managing this technologically complex and cumbersome project, and Leslie Groves — the project's overall director — regarded him with suspicion. Even before the war's end, Urey became deeply disenchanted with working for the military.

After the war, Urey used his laureate status to voice alarm about the prospect of nuclear warfare. He backed international control through world government as a way to control the military future of atomic energy. This was not a radical view in 1946; it was advanced in the US government's Report on the International Control of Atomic Energy, much of which had been drafted by Oppenheimer.

However, when the Soviet Union refused this plan for international control, which preserved the US atomic monopoly, advocates of world government found their loyalty as citizens questioned. In 1946, Urey was attacked by J. Parnell Thomas (who would go on to head the House Un-American Activities Committee) for being "one-world-minded", and not sufficiently patriotic. The FBI also investigated Urey, claiming that he belonged to several communist front organizations.



AFTER THE WAR, UREY USED HIS LAUREATE STATUS TO **VOICE ALARM** ABOUT THE PROSPECT OF NUCLEAR **WARFARE.**



**The Life and Science of Harold C. Urey**
MATTHEW SHINDELL
*University of Chicago Press (2019)*

Over this harrowing period, Urey lost faith in the ability of modern secular society to manage the new threats of the atomic age. Although he had long abandoned his parents' religion, he began to argue that Judaeo-Christianity was key to democracy. He attributed the success of science itself, with its commitments to honesty and credit, to religious ethics.

In the late 1940s, Urey used his expertise in mass spectrometry to begin work in geochemistry, and then in planetary science. It was a way to escape the orbit of the nuclear weapons establishment (although he still advised the US Atomic Energy Commission). With chemistry graduate Stanley Miller, he tested hypotheses on the origins of life by Soviet biochemist Alexander Oparin and biologist J. B. S. Haldane, and successfully produced amino acids by sparking a solution of water, methane, ammonia and hydrogen. In 1952, Urey published *The Planets*, a chemical treatise on the formation of the Solar System.

## LUNAR QUEST

Urey became influential during the early days of NASA, formed after the 1957 launch of the Soviet satellite Sputnik, offering the agency persuasive reasons to prioritize exploration of the Moon over other bodies. In 1969, he analysed lunar rocks collected during the Apollo 11 mission, which supported his theory of the Moon's common origin with Earth. He wanted the well-funded agency to test theories about the origins of the Solar System — experimentation beyond the reach of individual university scientists. Despite his influence, he was disappointed in this: NASA focused on crewed space exploration over questions of cosmogony. This last, frustrating chapter of Urey's life sheds light on the politics of mission-oriented research, in which popular interest or government priorities can take precedence over scientific questions.

Shindell keeps a tight focus on his biographical subject throughout the book. At times, the reader might wish that he had panned out a little more, to sketch the landscape of US cultural life in Urey's era, or comment on how the space race fitted within the global cold war. But this fine biography wonderfully shows how Urey's scientific contributions led chemistry in new directions, including to the Moon — and, in depicting the life of a leading scientist, Shindell probes the complex interplay of faith, values and politics in the United States. ■

---

**Angela N. H. Creager** *is Thomas M. Siebel Professor in the History of Science at Princeton University in New Jersey. Her latest book is* Life Atomic: A History of Radioisotopes in Science and Medicine.
*e-mail: creager@princeton.edu*

# Correspondence

## Students having to lead is shameful

The efforts of Amy Orben and other young researchers to fight the perverse incentives that dominate science right now are all the more impressive because these scientists are at the most vulnerable point of their careers (*Nature* **573**, 465; 2019). And, just as it's shameful that teenagers have to lead international action against the climate crisis, students and new postdocs should not have to spearhead these efforts.

Top-down pressure to improve research practice is needed. In my experience, even the smallest mandate from funders, publishers or performance assessors boosts incentive. Evaluators of research quality should openly declare how they measure a study's rigour and how that rigour contributes to quality scores. And funders should insist that institutions sign the San Francisco Declaration on Research Assessment, commonly known as DORA.

For their part, prestigious journals need to be more willing to accept registered reports and direct replications of studies they have published. And publication of the code and syntax behind analyses in manuscripts should be obligatory across all journals.

These are not radical proposals. Any one of them would reinforce trustworthiness in science.

**Jessica Butler** *University of Aberdeen, UK.*
*jessicabutler@abdn.ac.uk*

## More injustice for small island states

A statistical loophole is stopping research funding from getting to small island states — among the most vulnerable to climate change and among the least culpable for it. By contrast, international aid after disasters such as that caused by Hurricane Dorian (go.nature.com/33eggxt) can be swift and forthcoming.

In general, funders that provide aid and development use gross national income per capita as the sole measure of a country's development (go.nature.com/35djbd8). Countries with small populations, including the small island states, are therefore unlikely to appear on the list of nations eligible to receive aid, compiled for "statistical purposes" by the Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD).

Although the OECD specifically states that the DAC list is "not designed as guidance for aid or other preferential treatment", several UK research funders, for example, do use the DAC list to determine the eligibility of countries in research partnerships (go.nature.com/3jq92mm).

The solution is to include the official United Nations list of small island developing states (go.nature.com/2ab2xhf) as eligible partners in research and development programmes.

**Nicholas Higgs** *Cape Eleuthera Institute, Rock Sound, The Bahamas.*
*nickhiggs@ceibahamas.org*

## A first for African neuroscience

A new institute for neuroscience, the first of its kind in Africa, has opened at South Africa's University of Cape Town (go.nature.com/2rmjjpc). Its aim is to meet one neglected aspect of the projected health challenge of a growing population, namely illnesses of the brain and mind. The entire continent has just nine mental-health professionals for every 100,000 people (go.nature.com/35n1afj).

The institute will focus on conditions that are of the highest priority for Africa. One example is the surge in the number of children with brain infections caused by HIV and tuberculosis. And an improved understanding of brain development and health should help to inform strategies to maximize Africa's economic potential.

Cape Town's Neuroscience Institute promises to be an African centre of excellence, supported by scientists and clinicians from different fields. It will help to spread advances in the neurosciences across sub-Saharan Africa by acting as a nexus for training and collaboration through established networks such as the African Academy of Sciences (see also J. M. Wilmshurst *et al. Pediatrics* **137**, e20152741; 2016).

**A. Graham Fieggen*** *University of Cape Town, South Africa.*
*graham.fieggen@uct.ac.za*
**On behalf of 4 correspondents; see go.nature.com/2nyzbsu.*

## A research dream is crushed by politics

Yachay Tech, in Urcuquí, was founded as Ecuador's first research-intensive university in 2012 (go.nature.com/2mmqbpr). Once seen as an ambitious means of advancing a small emerging country, the university now seems to have followed a trend pervasive in Latin America: it has become primarily a teaching institution.

I was appointed as the university's vice-chancellor for research and innovation in 2016. I can attest to the hope of Ecuadorians from all walks of life that Yachay Tech would better their country's economy, and the lives of its people, through state-of-the-art higher education informed by research. Students were enthralled by the prospect of learning at a research institution of international calibre, and of engaging in research themselves (see C. Castillo-Chavez *et al. Science* **357**, 881; 2017).

This optimism has been beaten down by the headwinds of political meddling. Fractious partisan politics and financial constraints are compounded by a conspicuous lack of support from other important institutions and ministries. The incompetence of planners and builders for the City of Knowledge Yachay, where Yachay Tech is sited, is adding to the frustration.

I call on academics, politicians and concerned citizens to lend their weight to the development of research universities in emerging countries. These institutions need autonomy, national financial support, commitment to world standards of excellence and ethics — and freedom from political interference.

**Juergen Reichardt** *Australian Institute of Tropical Health and Medicine, James Cook University, Smithfield, Australia.*
*juergen.reichardt@jcu.edu.au*

## Citation doping not for Italy's elites

Factors other than citation doping could have contributed to the recent rise in the number of Italians among the 100,000 most highly cited scientists (see *Nature* http://doi.org/dcgj; 2019).

Of the 100,000 most highly cited scientists in the database compiled by John Ioannidis *et al.* (*PLoS Biol.* http://doi.org/gf6ckr; 2019), including some 2,000 Italians, we found that the proportion using self-citation to boost their research impact was probably only 2% (see P. D'Antuono and M. Ciavarella Preprint at https://arxiv.org/abs/1910.02948v1; 2019). The practice seems to be more common among early-career scientists who are otherwise less frequently cited.

ANVUR, the Italian agency for research evaluation (go.nature.com/2kwu5jj), should in our view exclude self-citations from future evaluations, to avoid this 'noise'.

We consider that the jump in the number of Italians in the 100,000 most highly cited researchers is a symptom of the overall health of the Italian research system. It underscores the positive effect of introducing ANVUR in 2006.

**Pietro D'Antuono, Michele Ciavarella** *Polytechnic University of Bari, Italy.*
*mciava@poliba.it*

NEUROSCIENCE

# The smoke clears over diabetes

**The discovery of a signalling axis that connects nicotine responses in the brain with glucose metabolism by the pancreas sheds light on why cigarette smoking increases the risk of diabetes.** SEE ARTICLE P.372

GIUSEPPE BRUSCHETTA & SABRINA DIANO

One of the many dangers of smoking is an increased risk of diabetes, because nicotine uptake leads to altered glucose metabolism and increased blood sugar levels[1–4]. A group of neurons activated by nicotine is located in the brain's medial habenula (mHb) region; they are characterized by expression of nicotine acetylcholine receptor (nAChR) proteins[5]. These neurons promote aversive responses to nicotine, but until now, no one knew whether they also control the diabetes-associated effects of smoking — and if so, how. On page 372, Duncan *et al.*[6] identify a signalling pathway that links nAChR-expressing neurons and blood-glucose regulation by the pancreas. It involves a diabetes-associated transcription factor, TCF7L2.

TCF7L2 is part of a signalling pathway[7] that regulates the production and secretion of the hormone glucagon-like peptide 1 (GLP-1). GLP-1 promotes insulin release from the pancreas, and insulin promotes glucose uptake from the blood into tissues to be stored as fat. Thus, changes in TCF7L2 expression or activity can influence glucose metabolism. (Of note, the pathway is complex because activation of GLP-1 receptors by GLP-1 in turn leads to TCF7L2 activity, creating a signalling loop[7].)

Duncan *et al.* discovered that TCF7L2 is expressed in the mHb. Given the role of the mHb in the body's response to nicotine, the authors set out to investigate the possibility that TCF7L2 expression in this region is involved in nicotine-induced alterations in glucose metabolism. The researchers first generated two strains of mice — one in which the function of TCF7L2 was impaired in all tissues, and one in which the protein's expression could be inhibited in the mHb by introducing a virus carrying a short RNA sequence. They then provided the mice with a nicotine solution through a drip, rigged such that the animals could push a lever to receive an injection of nicotine at will. Both mice strains showed greater nicotine intake than did control animals.

Prolonged stimulation by nicotine causes nAChR-expressing neurons to become desensitized and stop responding to the molecule[8]. Further examination of the mutant animals indicated that TCF7L2 improves the ability of nAChR-expressing neurons to recover from this desensitization, presumably therefore promoting aversive responses to smoking. TCF7L2 seems to exert this effect by promoting signalling through cyclic AMP (ref. 9) — a molecule often involved in intracellular signal transduction.

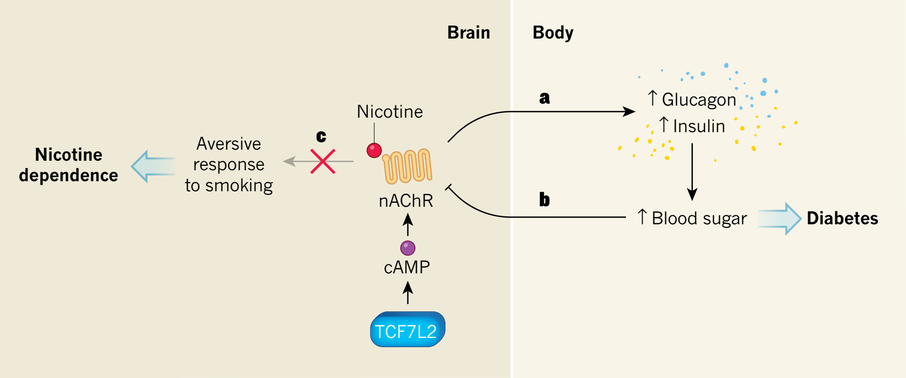


**Figure 1 | A link between nicotine, diabetes and addiction.** Nicotine activates nicotine acetylcholine receptor (nAChR) proteins in neurons. Activation of a population of nAChR- expressing neurons in the brain's medial habenula (mHb) region leads to aversive responses to nicotine. **a**, Duncan *et al.*[6] report that activation of these receptors in mice also leads to increases in the release of glucagon and insulin hormones from the pancreas. This leads to increased blood sugar levels — a change associated with an increased risk of diabetes in humans. **b**, The authors find that increased blood sugar levels in turn inhibit the activity of the nAChR-expressing neurons, establishing a feedback loop. **c**, As a result of this feedback, aversive responses to smoking are no longer triggered, resulting in nicotine dependence. The entire circuit is modulated by the protein TCF7L2 that, acting through a second messenger molecule called cAMP, mediates the sensitivity of nAChR-expressing neurons to nicotine.

Duncan *et al.* next demonstrated that TCF7L2 augments the ability of nicotine to increase blood-sugar levels: inhibition of TCF7L2 or GLP-1 receptors abolished this effect of nicotine, whereas a drug that stimulates GLP-1 receptors had the opposite effect. Furthermore, when the authors subjected wild-type mice to nicotine exposure and subsequent withdrawal, they observed increased levels of glucose, glucagon and insulin in the blood — indicative of an inability to properly regulate glucose metabolism. The effects were lessened in the mutant mice. Together, these analyses suggest that TCF7L2 activity decreases dependence on nicotine, but, conversely, hampers the body's ability to counteract the detrimental effect of nicotine on glucose metabolism.

To explore the neural circuits involved, Duncan and colleagues injected wild-type mice with a fluorescently labelled 'retrograde' virus, which travels along any neurons connected to an injection site. After injection in the pancreas, the authors found the virus in several areas of the brain, including the mHb, indicating the existence of a neuronal signalling axis from the mHb to the pancreas, by way of other brain regions. Thus nicotine, by stimulating mHb neurons in a TCF7L2-dependent manner, engages neuronal signalling to the pancreas and so alters blood glucose levels.

A 2011 report indicated that only around 6% of smokers quit successfully each year[10], and data indicate that people with diabetes find it harder to quit than do people without the condition[11]. In a final set of experiments, Duncan *et al.* gave mice sucrose every day for 6 weeks and showed that the subsequent increase in blood sugar led to reduced TCF7L2 levels and a decrease in nicotine-evoked activity in nAChR-expressing neurons. This observation suggests a feedback mechanism that could explain why it is harder for people with diabetes to give up smoking[11] (Fig. 1).

Duncan and colleagues' study sheds new light on the role of nicotine in the regulation of glucose metabolism. However, there are some caveats to consider. TCF7L2 is expressed not only in the mHb, but also in other brain regions[12]. The pancreas, fat tissues and intestine — all of which are involved in glucose metabolism — also express TCF7L2. The possible contribution of these structures to nicotine-induced dysregulation of glucose metabolism needs to be considered and evaluated.

Furthermore, nicotine is a strong activator of the hypothalamus–pituitary–adrenal (HPA) axis — a network that promotes the release of stress hormones[13]. The function of the mHb is altered by stress, and stress hormones induce changes in blood glucose[13]. As such, it is conceivable that the some of the effects reported by the authors reflect not only the direct effect of nicotine on the mHb, but also an indirect effect through HPA-axis activation.

An interesting question is whether the effect of nicotine on the mHb–pancreas axis is different in males and females. In support of this idea, more men than women smoke, and the risk of female smokers developing diabetes is much greater than the risk for male smokers, compared with their non-smoking counterparts[4]. Furthermore, nicotine withdrawal induces greater weight gain in women than in men[14].

More broadly, further investigation is needed to confirm the role of the mHb–pancreas circuit in humans. Tobacco addiction in humans involves the interplay of pharmacological, genetic, social and environmental factors. Therefore, the full picture of nicotine's role in diabetes is likely to involve much more than a single regulatory circuit. Finally, this work raises the question of whether and how TCF7L2 could be targeted to combat tobacco dependency and diabetes. The feasibility of this tantalizing idea will require much more investigation in both mice and humans. ■

**Giuseppe Bruschetta** *and* **Sabrina Diano** *are in the Department of Cellular and Molecular Physiology and the Program of Integrative Cell Signaling and Neurobiology of Metabolism, Yale University School of Medicine, New Haven, Connecticut 06520, USA.*
*e-mail: sabrina.diano@yale.edu*

1. Chiolero, A., Faeh, D., Paccaud, F. & Cornuz, J. *Am. J. Clin. Nutr.* **87**, 801–809 (2008).
2. Chase, H. P. *et al. J. Am. Med. Assoc.* **265**, 614–617 (1991).
3. Madsbad, S. *et al. Diabetes Care* **3**, 41–43 (1980).
4. Will, J. C., Galuska, D. A., Ford, E. S., Mokdad, A. & Calle, E. E. *Int. J. Epidemiol.* **30**, 540–546 (2001).
5. Benowitz, N. L. *N. Engl. J. Med.* **362**, 2295–2303 (2010).
6. Duncan, A. *et al. Nature* **574**, 372–377 (2019).
7. Liu, Z. & Habener, J. F. *J. Biol. Chem.* **283**, 8723–8735 (2008).
8. Picciotto, M. R., Addy, N. A., Mineur, Y. S. & Brunzell, D. H. *Prog. Neurobiol.* **84**, 329–342 (2008).
9. Khiroug, L., Sokolova, E., Giniatullin, R., Afzalov, R. & Nistri, A. *J. Neurosci.* **18**, 2458–2466 (1998).
10. Centers for Disease Control and Prevention. *MMWR Morbid. Mortal. Wkly Rep.* **60**, 1513–1519 (2011).
11. Solberg, L. I., Desai, J. R., O'Connor, P. J., Bishop, D. B. & Devlin, H. M. *Ann. Fam. Med.* **2**, 26–32 (2004).
12. Lee, S., Lee, C. E., Elias, C. F. & Elmquist, J. K. *J. Comp. Neurol.* **517**, 925–939 (2009).
13. Tweed, J. O., Hsia, S. H., Lutfy, K. & Friedman, T. C. *Trends Endocrinol. Metab.* **23**, 334–342 (2012).
14. Pirie, P. L., Murray, D. M. & Luepker, R. V. *Am. J. Public Health* **81**, 324–327 (1991).

---

CANCER

# Brain tumours reset their clocks

**The body's circadian clock ensures the rhythmic expression of some genes across the day. The catalogue of genes under circadian control changes in an aggressive brain cancer — a discovery that might open up a new avenue for treatment.**

**GUIOMAR SOLANAS & SALVADOR A. BENITAH**

All organisms have an internal circadian clock, which ensures that physiological functions occur at the right time of day — for instance, that the intestine, pancreas and liver are ready to metabolize food when you eat rather than when you are sleeping. Writing in *Cancer Discovery*, Dong *et al.*[1] report that the cells responsible for initiating a specific type of aggressive brain tumour, glioblastoma, rely on an altered circadian clock to grow. What's more, drug-based inhibition of the cells' molecular clock can kill them.

Glioblastomas are the most prevalent and aggressive tumour of the central nervous system. Fewer than 6% of patients survive for five years after diagnosis[2]. Cells in a glioblastoma often have varied gene-expression profiles. This, coupled with the fact that glioblastoma-initiating stem cells (GSCs) act to maintain the tumour, means that glioblastomas can rapidly develop resistance to conventional therapies[3]. New treatments are therefore urgently needed.

Disruption of the circadian clock, either because of lifestyle choices or because of mutations in core clock genes, is associated with a higher incidence of tumours[4]. In some tissues, clock genes can be co-opted to promote cancer (they are said to act as oncogenes), whereas in others they act as tumour suppressors[5–7]. The origin of such differences is an open question that, when answered, will help researchers to identify the mechanisms by which tumour cells hijack the molecular clock machinery to increase their chances of survival.

Dong and colleagues show that two key clock genes, *BMAL1* and *CLOCK*, are co-opted to act as oncogenes in glioblastoma. The authors first observed that the genes are essential for the survival and proliferation of GSCs *in vitro*. By contrast, neither differentiated glioblastoma cells nor normal neural stem cells (from which GSCs arise[8]) seem to depend on the genes in this way. The authors validated these findings by showing a strong correlation

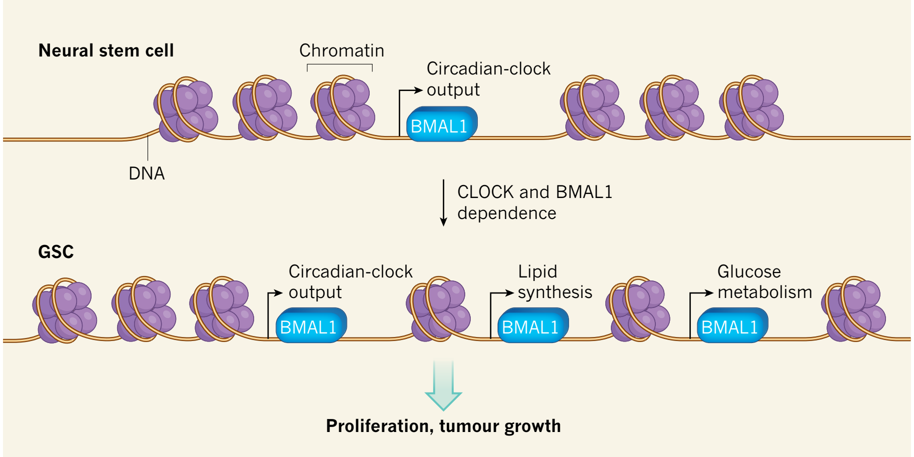


**Figure 1 | Circadian reprogramming in cancer stem cells.** The proteins BMAL1 and CLOCK are core components of the body's circadian clock. In neural stem cells, the proteins bind to specific regions of DNA (which is packaged around proteins as chromatin) to promote expression of the circadian-clock output — a collection of genes that are expressed in oscillating rhythms across the day (only BMAL1 is shown here). Dong *et al.*[1] report that, when neural stem cells become cancerous glioblastoma-initiating stem cells (GSCs), they become dependent on BMAL1 and CLOCK for survival. Changes in chromatin packaging enable BMAL1 and CLOCK to bind to more sites across DNA, promoting the expression of genes involved in lipid synthesis and glucose metabolism. Activation of these two metabolic pathways promotes GSC proliferation and so tumour growth.

between the expression of some of the core clock components and patient outcomes.

The researchers went on to show that a process called circadian reprogramming might explain why GSCs depend on the circadian clock. Circadian reprogramming involves changes in the circadian-clock output — that is, in the collection of genes in a given cell or tissue that are under the control of the clock, and so are expressed in oscillating rhythms across the day. Dong *et al.* demonstrated that the circadian-clock output of GSCs includes genes involved in glucose metabolism and lipid synthesis, whereas the circadian-clock output of normal neural stem cells does not. Changes in glucose metabolism and lipid synthesis have been previously shown to aid cancer progression[9].

In addition, Dong and colleagues observed that the metabolic capacity of GSCs changed in the absence of *BMAL1* and *CLOCK*. The group showed that circadian reprogramming in GSCs is mediated by changes in chromatin — the DNA–protein complex in which DNA is packaged. More regions of chromatin are open in GSCs than in normal neural stem cells, allowing the BMAL1 and CLOCK proteins to bind to and activate different genes. The authors then linked these data by showing that BMAL1 and CLOCK regulate the expression of genes involved in lipid metabolism in GSCs, indicating that the oncogenic activity of the clock genes might involve metabolic pathways (Fig. 1).

Previous reports have described circadian reprogramming in response to various stimuli, such as changes in diet, physiological ageing or exercise[10–13]. In all these cases, circadian reprogramming is a fast and effective way to respond to changing external demands. Circadian reprogramming has also been observed between organs — for instance, reprogramming in the livers of mice that have developed lung cancer probably ensures that the liver provides sufficient energy for the tumour cells to grow efficiently[14]. The picture that is emerging is of circadian reprogramming as a common mechanism to help cells, tissues and whole organisms adapt to change, whether they are healthy or cancerous.

In a final set of experiments, Dong *et al.* showed that small molecules that repress *BMAL1*, either directly or indirectly, strongly inhibit the self-renewing potential of GSCs. Mice that carried GSCs from patients survived longer if they were treated with one of these molecules than they did without treatment.

Caution is needed when considering translating these findings to humans, because the small molecules used by Dong and colleagues also affect the activity of the clock machinery systemically, potentially perturbing normal physiological processes in healthy tissues — this might induce damage accumulation and signs of premature ageing[15]. A better alternative might be to target the factors that induce circadian reprogramming in GSCs. Such an approach should block circadian-related changes in gene expression in cancer cells without perturbing the clock in the rest of the organism.

What might these factors be? There is likely to be a mixture, some intrinsic to the cells, others extrinsic, probably acting synergistically. For example, as in the current study, a change in energy requirements when a cell becomes cancerous can lead to changes in the metabolic products generated in that cell; this, in turn, can affect chromatin remodelling, changing the catalogue of genes available to be activated and thereby altering the rhythmic transcription of genes[16]. Outside the cell, signalling pathways involving the hormone insulin and the neurotransmitter molecule adrenaline are both altered in tumours, and can re-entrain the cancer-cell clock, thus integrating whole-body information into the cell's circadian output[15].

These systemic pathways might represent therapeutic targets to treat cancer. However, the complex effect of these pathways on circadian reprogramming in cancer cells is still poorly understood. Nonetheless, Dong *et al.* have opened a new chapter in the search for therapeutic targets for aggressive and incurable glioblastomas. ■

**Guiomar Solanas** *and* **Salvador A. Benitah** *are at the Institute for Research in Biomedicine Barcelona, Barcelona Institute of Science and Technology, 08028 Barcelona, Spain.* **S.A.B.** *is also at the Catalan Institution for Research and Advanced Studies (ICREA).*
*e-mails: guiomar.solanas@irbbarcelona.org; salvador.aznar-benitah@irbbarcelona.org*

1. Dong, Z. *et al. Cancer Discov.* https://doi.org/10.1158/2159-8290.CD-19-0215 (2019).
2. Bohn, A. *et al. PLoS ONE* **13**, e0198581 (2018).
3. Weller, M. *et al. Nature Rev. Dis. Primers* **1**, 15017 (2015).
4. Verlande, A. & Masri, S. *Trends Endocrinol Metab.* **30**, 445–458 (2019).
5. Janich, P. *et al. Nature* **480**, 209–214 (2011).
6. Puram, R. V. *et al. Cell* **165**, 303–316 (2016).
7. Fekry, B. *et al. Nature Commun.* **9**, 4349 (2018).
8. Lee, J. H. *et al. Nature* **560**, 243–247 (2018).
9. Pascual, G., Dominguez, D. & Benitah, S. A. *Dis. Model Mech.* **11**, dmm032920 (2018).
10. Sato, S. *et al. Cell* **170**, 664–677 (2017).
11. Solanas, G. *et al. Cell* **170**, 678–692 (2017).
12. Sato, S. *et al. Cell Metab.* **30**, 92–110 (2019).
13. Eckel-Mahan, K. L. *et al. Cell* **155**, 1464–1478 (2013).
14. Masri, S. *et al. Cell* **165**, 896–909 (2016).
15. Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V. & Antoch, M. P. *Genes Dev.* **20**, 1868–1873 (2006).
16. Mohawk, J. A., Green, C. B. & Takahashi, J. S. *Annu. Rev. Neurosci.* **35**, 445–462 (2012).

---

AGEING

# Neural excitation moderates lifespan

**Signals emanating from the nervous system are potent modulators of longevity. It now seems that overall neural excitation is also a key determinant of lifespan. SEE ARTICLE P.359**

NEKTARIOS TAVERNARAKIS

The question of why and how we age — and why only a minority of humans live to become centenarians — has fascinated people for millennia. Over the past few decades, we have learnt that the rate of ageing is highly sensitive to intrinsic and extrinsic cues, and that these cues act, by means of numerous genetic pathways, to regulate the cellular and systemic processes that ultimately influence ageing[1].

On page 359, Zullo and colleagues[2] uncover a new twist in the saga: an unexpected link between the nervous system and ageing. They show that overall neuronal excitation is a major determinant of lifespan, and that it is higher in short-lived individuals and lower in the long-lived. The authors also characterize some of the molecular players in this effect, and tie it to a well-known regulator of lifespan: signalling by the hormone insulin or by insulin-like growth factor 1 (IGF1).

Ageing affects the nervous system in a complex way that is not yet fully understood[3–5]. Perhaps less intuitively, this relationship also works in the opposite direction: signals from the nervous system can modulate the rate of ageing of the whole organism[6–10]. But although the nervous system is known to influence longevity in species ranging from invertebrates to mammals, the underlying molecular mechanisms have been unclear.

Zullo and colleagues began their investigation by studying brain tissue from aged humans who had shown no cognitive deficits before their death. The authors analysed gene-expression profiles from the frontal cortex, and uncovered an intriguing correlation: genes involved in neural excitation and in the function of the synaptic connections between neurons are downregulated in long-lived individuals, but genes required for inhibitory neurotransmission are not.

How might this occur? The authors found that the downregulated genes are probably targets of the transcriptional regulator protein REST — a general repressor of genes involved in neuronal excitation and synaptic function[11]. Previous studies[12,13] had implicated REST in preventing hyperexcitation of the neuronal network, maintaining its steady state, resisting oxidative stress and protecting neurons over time. (For instance, deleting the *Rest* gene increases neural activity in the mouse cortex and renders animals vulnerable to blockers of inhibitory neurotransmission, further exacerbating neural excitation and triggering epilepsy.) The new findings directly associate long human lifespan with increased REST activity and reduced neural excitation.

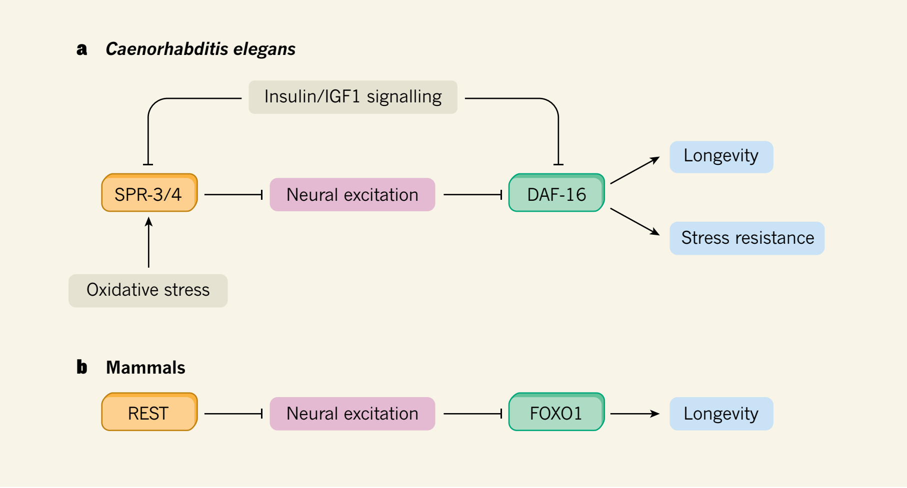


**Figure 1 | Lifespan is regulated by neural excitation in worms and mammals.** **a**, Zullo and colleagues[2] show that, in the nematode worm *Caenorhabditis elegans*, the proteins SPR-3 and SPR-4 reduce the expression of genes involved in neural excitation and transmission across synapses. SPR-3 and SPR-4 therefore act to quench neural excitation. Subsequent activation of the transcription factor DAF-16 (normally inhibited by neural excitation) promotes longevity and resistance to oxidative stress. Oxidative stress and signalling through insulin or insulin-like growth factor 1 (IGF1) are both known to affect lifespan. The new findings suggest that they do this, in part, through their effects on SPR-3/4 and neural excitation. **b**, The authors also find that, in humans and mice, the SPR-3 and SPR-4 counterpart REST downregulates genes involved in neural excitation in the brain's cortex. The ensuing tempering of neural excitation activates the DAF-16 counterpart FOXO1. REST expression is increased in the cortex of long-lived humans, whereas genes involved in neural excitation are downregulated.

Is this association merely a corollary of the ageing process, or is there a causal relationship? To find out, Zullo and colleagues turned to the nematode worm *Caenorhabditis elegans* — a malleable test bed that has been invaluable in unpicking the mechanisms that modulate lifespan[14]. The authors found that neural activity increases as the worm ages. In addition, interventions that inhibit either overall neural excitation and synaptic neurotransmission or signalling by neuropeptide molecules extend the lifespan of *C. elegans*. In effect, tempering excitatory neurotransmission to reduce overall neural activity is enough to make worms longer-lived. By contrast, suppressing inhibitory neurotransmission increases neural activity and shortens lifespan. Overall neural excitation is, therefore, an important regulator of lifespan in worms and humans.

Digging deeper into this process in worms, the authors focused on the SPR-3 and SPR-4 proteins[15], which are counterparts of mammalian REST. This is where the research began to reveal links with insulin/IGF1 signalling, which is a key part of the cellular response to the presence of nutrients. Low insulin/IGF1 signalling is associated with long lifespan in worms.

Zullo *et al.* found that the longevity conferred by reduced neural activity requires DAF-16, a transcription factor that is also needed for the extended lifespan linked to low insulin/IGF1 signalling in *C. elegans*. Moreover, neuronal SPR-3 and SPR-4 are key to the increase in lifespan seen under conditions of low insulin/IGF1 signalling. Genes required for neuronal excitation are downregulated by low insulin/IGF1 signalling in an SPR-3/4-dependent manner. In addition, worms carrying mutations in the insulin receptor DAF-2 show reduced neural excitation that is instigated by SPR-3 and SPR-4 and is required for activation of DAF-16. Similarly, SPR-3 and SPR-4 are needed to activate DAF-16 under conditions of oxidative stress. Of note, SPR-3/4 depletion restores higher levels of neural excitation in animals carrying DAF-2 mutations, compromising their exceptional longevity.

Collectively, these findings in *C. elegans* indicate that stress and insulin/IGF1 signalling converge on SPR-3 and SPR-4 to modulate neural activity. In turn, this influences DAF-16, another point of convergence that integrates neural excitation and insulin/IGF1 signals to promote stress tolerance and longevity (Fig. 1a). Exactly how DAF-16 is activated by reduced neural excitation remains to be seen.

A similar signal-transduction pathway seems to be at work in mammals (Fig. 1b). Zullo *et al.* find that, in humans, the expression and levels of REST in cell nuclei correlate with those of the DAF-16 counterpart FOXO1. Moreover, both REST and FOXO1 are found in neurons in the human prefrontal cortex. The authors showed that repressing neural excitation in mouse cortical neurons grown in culture increases the expression and nuclear levels of FOXO1. And the age-dependent rise in nuclear FOXO1 in mice requires REST. The parallels between nematodes and mammals suggest that the REST–FOXO1 (or the SPR-3/4–DAF-16) axis is a key part of the mechanisms by which nervous system function influences ageing. Moreover, a reduction in overall neural excitation is a major contributor to the lifespan extension caused by low insulin/IGF1 signalling.

***"The new findings have revealed a previously unappreciated conduit for integrating neural activity and metabolism."***

These findings shine new light on previous work. For example, certain anticonvulsant drugs have been found to promote longevity in *C. elegans*[16] — again implicating overall neural activity in regulating lifespan. However, these compounds act in complex ways, and their anti-ageing effects might not be entirely reliant on the nervous system. Moreover, unlike REST-mediated quenching of neural excitation, some anticonvulsants function independently of DAF-16, and further extend the lifespan of animals with DAF-2 mutations.

Another line of work, again in *C. elegans*, has offered a link between longevity and the inhibition of signalling by the neurotransmitter molecule serotonin, which is involved in the organismal response to nutrients[17]. That research showed that antidepressants that block serotonin receptors extend lifespan, probably by simulating dietary restriction (known to promote longevity universally). Given that dietary restriction is associated with low insulin/IGF1 signalling, which limits overall neural excitation through REST, could REST (or SPR-3/4) also contribute to lifespan extension under nutritional stress? Zullo and colleagues consider this unlikely, because inhibiting neural excitation in *C. elegans* late in adulthood — even after the worms have stopped feeding — still extends lifespan. But some role for this molecular axis is possible, given that cessation of food intake is not always equivalent to a state of perceived calorie restriction.

As well as offering insight into the link between overall neural excitation and ageing, Zullo and colleagues' findings provide a previously unappreciated conduit for integrating neural activity and metabolism, through the insulin/IGF1 pathway. This integration could fine-tune an organism's physiology and orchestrate appropriate behavioural adaptations towards optimizing fitness and enhancing survival. What is more, by buffering changes in overall neural excitation and maintaining a proper balance in neuronal-network activity, REST might also prevent age-related neurological disorders to boost longevity in humans. Indeed, accumulating evidence couples neural overexcitation to Alzheimer's disease[18–20]. So REST and other molecules that control neural excitability are possible targets for interventions aimed at battling the decline and maladies of old age. ■

**Nektarios Tavernarakis** *is at the Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, and the School of Medicine, University of Crete, Heraklion 71110, Crete, Greece.*
*e-mail: tavernarakis@imbb.forth.gr*

1. Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. *Cell* **153**, 1194–1217 (2013).
2. Zullo, J. M. *et al. Nature* **574**, 359–364 (2019).
3. Lu, A. T. *et al. Nature Commun.* **8**, 15353 (2017).
4. Camandola, S. & Mattson, M. P. *EMBO J.* **36**, 1474–1492 (2017).
5. Mattson, M. P. & Arumugam, T. V. *Cell Metab.* **27**, 1176–1199 (2018).
6. Ailion, M., Inoue, T., Weaver, C. I., Holdcraft, R. W. & Thomas, J. H. *Proc. Natl Acad. Sci. USA* **96**, 7394–7397 (1999).
7. Apfeld, J. & Kenyon, C. *Nature* **402**, 804–809 (1999).
8. Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. *Science* **290**, 147–150 (2000).
9. Alcedo, J. & Kenyon, C. *Neuron* **41**, 45–55 (2004).
10. Bishop, N. A., Lu, T. & Yankner, B. A. *Nature* **464**, 529–535 (2010).
11. Schoenherr, C. J. & Anderson, D. J. *Science* **267**, 1360–1363 (1995).
12. Pozzi, D. *et al. EMBO J.* **32**, 2994–3007 (2013).
13. Lu, T. *et al. Nature* **507**, 448–454 (2014).
14. Kenyon, C. J. *Nature* **464**, 504–512 (2010).
15. Smialowska, A. & Baumeister, R. *Neurodegen. Dis.* **3**, 227–232 (2006).
16. Evason, K., Huang, C., Yamben, I., Covey, D. F. & Kornfeld, K. *Science* **307**, 258–262 (2005).
17. Petrascheck, M., Ye, X. & Buck, L. B. *Nature* **450**, 553–556 (2007).
18. Siskova, Z. *et al. Neuron* **84**, 1023–1033 (2014).
19. Ping, Y. *et al. PLoS Genet.* **11**, e1005025 (2015).
20. Frazzini, V. *et al. Cell Death Dis.* **7**, e2100 (2016).

---

INFECTIOUS DISEASE

# Malaria mosquitoes go with the flow

**The rapid return of mosquitoes to African semi-desert regions when the dry season ends was an unsolved mystery. A surprising solution to the puzzle is the long-range migration of mosquitoes on high-altitude winds.** SEE LETTER P.404

### NORA J. BESANSKY

During the long dry season in the semi-desert region of Africa known as the Sahel, malaria transmission ceases because the mosquitoes that can transmit the disease (termed malaria mosquitoes or vectors) disappear, along with the surface water required for the development of the next generation of mosquitoes. Yet with the first rains that end the dry season, adult numbers surge more quickly than can be explained by resumed breeding in newly rain-filled sites. Evidence to explain this adult population boom has remained elusive for decades. On page 404, Huestis *et al.*[1] report high-altitude sampling of malaria vectors in the Sahel, which revealed data consistent with long-range wind-borne migration of mosquitoes.

Insect flight typically occurs close to the ground, in a habitat patch that provides all of the insect's essential resources such as food, shelter, mates and breeding sites. Among malaria vectors, this type of foraging flight rarely exceeds a distance of five kilometres[2]. By contrast, during long-distance migration, insects ascend to altitudes as high as 2–3 km, where fast air currents transport them downwind for hundreds of kilometres in a few hours[3]. This behaviour is beneficial[3] for insects moving in seasonally favourable directions.

The migration of monarch butterflies (*Danaus plexippus*) between North America and Mexico is one of the most widely known insect migrations, but the extent to which other insects engage in long-distance migration is under-appreciated, because these high-altitude flights are undetectable without technology such as radar. The type of radar that can detect larger insects (those heavier than 10 milligrams) had been mainly used to track just a few agricultural pests, until a 2016 study of the southern United Kingdom[4] used such radar to investigate insect migration in general. This study revealed that an estimated 16.5 billion insects migrate annually at high altitude (defined in this case as a height of more than 150 metres) above the $70,000\ \text{km}^2$ study area, indicating that wind-borne insect migration can occur on a strikingly large scale.

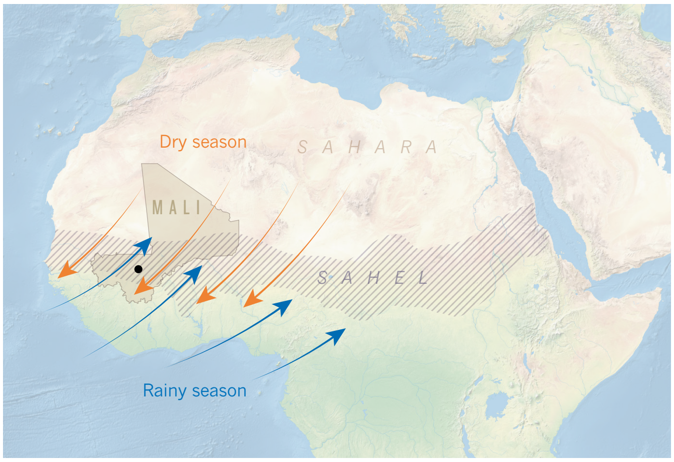


**Figure 1 | High-altitude winds enable the seasonal migration of African mosquitoes.** Huestis *et al.*[1] report that certain types of mosquito that can transmit malaria undergo long-distance wind-borne journeys. The authors studied sites in Mali (region marked with a black circle) in a semi-desert region of Africa called the Sahel. In the rainy season, there is a sudden rapid rise in the number of mosquitoes in the Sahel. The seasonal patterns of high-altitude wind directions (coloured arrows) are consistent with rainy-season winds transporting mosquitoes into the Sahel from southerly sites, where mosquitoes reside throughout the year. During the dry season, winds from the north blow into the Sahel, which could transport mosquitoes southwards.

Current radar technology does not detect small insects (lighter than 10 mg) such as mosquitoes, which must instead be tracked by sampling using aerial nets. In the UK study[4], such insect capture provided evidence that three trillion small insects undertake high-altitude migrations, a number that substantially exceeds that of the larger radar-tracked insects in the same area. These migrations, termed mass seasonal bioflows[4], involve representatives of all major insect orders[3], including Diptera, to which mosquitoes belong. Seasonal patterns in the direction of high-altitude winds can enable consistent routes for these bioflows (Fig. 1).

Huestis and colleagues studied four villages in the Sahel region of Mali. The possibility that wet-season mosquito populations are re-established there by adults flying from the nearest year-round populations was excluded in a previous study[5] by this team. This is because the distance of more than 150 km to such sites is prohibitively long for self-powered mosquito flight.

A second possibility is that mosquitoes maintain a local presence and survive during the dry season, hidden away in a state of dormancy termed aestivation. Important, albeit indirect, support for this hypothesis came from extensive population time-series analysis from that earlier study[5], which showed beyond reasonable doubt that a mosquito vector species called *Anopheles coluzzii* persists locally in the dry season in as-yet-undiscovered places. However, the data were not consistent with this outcome for other malaria vectors in the study area — the species *Anopheles gambiae* and *Anopheles arabiensis* — leaving wind-powered long-distance migration as the only remaining possibility to explain the data[5].

Both modelling[6] and genetic studies[7] support the idea of long-distance migration to explain the seasonal dynamics of malaria mosquitoes in the Sahel, but many researchers have instead long discounted this phenomenon as being rare, accidental and inconsequential. This entrenched attitude has been difficult to dispel given the challenge of obtaining compelling direct evidence.

Huestis *et al.* met this challenge through aerial sampling of insects using sticky nets tethered to helium-filled balloons stationed in the villages that they studied. Nets suspended at set altitudes ranging from 40 to 290 metres above ground were launched at night (malaria mosquitoes are nocturnal), for about 10 consecutive nights each month over a span of 22–32 months. During a total of 617 sampling nights, 461,100 insects were caught, which included 2,748 mosquitoes. Careful controls by the authors enabled them to conclude that the insects were captured at altitude and not during balloon deployment near the ground.

Among the mosquitoes captured were *A. gambiae* and *A. coluzzii*, as well as four other species of malaria vector. Comparable distributions of species across villages and years, and consistent peaks in insect captures in the mid to late rainy season, indicate that high-altitude migration of malaria vectors is deliberate rather than accidental. Moreover, the annual malaria vector bioflow predicted to cross a hypothetical 100-km line joining the authors' sampling sites exceeds 50 million insects, suggesting that high-altitude migration is common rather than rare. Simulated migratory trajectories for these vectors yield maximal distances of around 300 km, assuming one 9-hour high-altitude journey.

From this work and their previous study[5], Huestis and colleagues have finally resolved in broad outline the 'dry-season paradox' in favour of two non-mutually exclusive strategies: long-distance migration and local persistence. Yet many knowledge gaps remain.

Perhaps the most important of these is whether wind-borne migration includes malaria mosquitoes infected with malaria-causing parasites. The authors make much of the fact that female insects (only females transmit malaria) outnumber males by a ratio of more than 4:1 in the mosquitoes they captured, that more than 90% of the females had taken at least one blood meal before their flight, and that 31% of those meals were from humans, implying possible mosquito exposure to malaria parasites and the potential to spread infection over great distances.

However, the authors failed to detect parasite infections in their aerially sampled malaria vectors, a result that they assert is to be expected given the small sample size and the low parasite-infection rates typical of populations of malaria vectors. A problem with this argument is that the typical infection rates they mention are based on one specific mosquito body part (salivary glands), rather than the unknown but undoubtedly much higher infection rates that would be obtained if whole mosquito bodies were used to test for parasite infection. Further research will be required to flesh out this and many other fundamental issues raised by Huestis and colleagues' study.

If it is confirmed that there are wind-borne mosquitoes infected with the malaria-causing parasite, the implications of this would include the possibility of the reintroduction of disease into places where malaria has been previously eliminated, as well as the potential for the long-distance spread of drug-resistant parasites. Wind-borne malaria vectors, whether or not they are infected with parasites, could also profoundly affect the success of vector-control efforts. For example, migration could foster the long-distance spread of insecticide-resistant mosquitoes, worsening an already dire situation, given the current spread of insecticide resistance in mosquito populations. This would be a matter of great concern because insecticides are the best means of malaria control currently available[8]. However, long-distance migration could facilitate the desirable spread of mosquitoes for gene-based methods of malaria-vector control. One thing is certain, Huestis and colleagues have permanently transformed our understanding of African malaria vectors and what it will take to conquer malaria. ■

**Nora J. Besansky** *is in the Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA.*
*e-mail: nbesansk@nd.edu*

1. Huestis, D. L. *et al. Nature* **574**, 404–408 (2019).
2. Service, M. W. *J. Med. Entomol.* **34**, 579–588 (1997).
3. Chapman, J. W., Drake, V. A. & Reynolds, D. R. *Annu. Rev. Entomol.* **56**, 337–356 (2011).
4. Hu, G. *et al. Science* **354**, 1584–1587 (2016).
5. Dao, A. *et al. Nature* **516**, 387–390 (2014).
6. North, A. R. & Godfray, H. C. J. *Malar. J.* **17**, 140 (2018).
7. Lehmann, T. *et al. Evol. Appl.* **10**, 704–717 (2017).
8. Bhatt, S. *et al. Nature* **526**, 207–211 (2015).

---

QUANTUM PHYSICS

# Sounds of a supersolid

**Ultracold gases of dipolar atoms can exhibit fluid and crystalline oscillations at the same time, illuminating the ways in which different kinds of sound propagate in the quantum state of matter known as a supersolid.** SEE LETTERS P.382 & P.386

SEAN M. MOSSMAN

The quest to realize an exotic state of matter called a supersolid has intrigued researchers since initial theoretical predictions of its existence[1] were made in the late 1960s. A supersolid combines properties of a crystalline solid and a superfluid — a fluid that flows without resistance. After intense debate about a possible observation of supersolidity in solid helium[2], ultracold atomic gases have emerged as a powerful platform for investigating supersolid behaviour[3–5]. Tanzi *et al.*[6] and Guo *et al.*[7], on pages 382 and 386, respectively, and Natale *et al.*[8], writing in *Physical Review Letters*, have now made direct observations of supersolid dynamics. The teams have excited these exotic systems, and tuned in to the sounds of a supersolid for the first time.

To gain an intuitive picture of a supersolid, consider a narrow channel of fluid. Imagine that we turn a dial on our experiment and regularly spaced droplets begin to form — regions of high density that are connected through a background flow of liquid. The emerging droplets have a rigidity in that they tend to hold a fixed spacing, whereas the fluid that comprises them flows between the droplets without resistance. As we continue to turn the dial, this supersolid breaks as the droplets form more tightly until each one is isolated from its neighbours.

Researchers make such a state in the laboratory by using laser beams to suspend a collection of atoms inside a vacuum chamber. They then cool these atoms to some of the lowest temperatures in the Universe — about 50 nanokelvin. At these temperatures, the atoms condense into a single quantum state, a phase of matter known as a Bose–Einstein condensate (BEC). In such BECs, the atoms are superfluid and move in concert as a single quantum object.

The BECs produced for the three current experiments use atoms, such as erbium or dysprosium, that have strong permanent magnetic dipole moments. These atoms interact over long ranges, much as do the atoms in liquid helium[9], allowing for a roton — a kind of excitation that has a particular momentum.

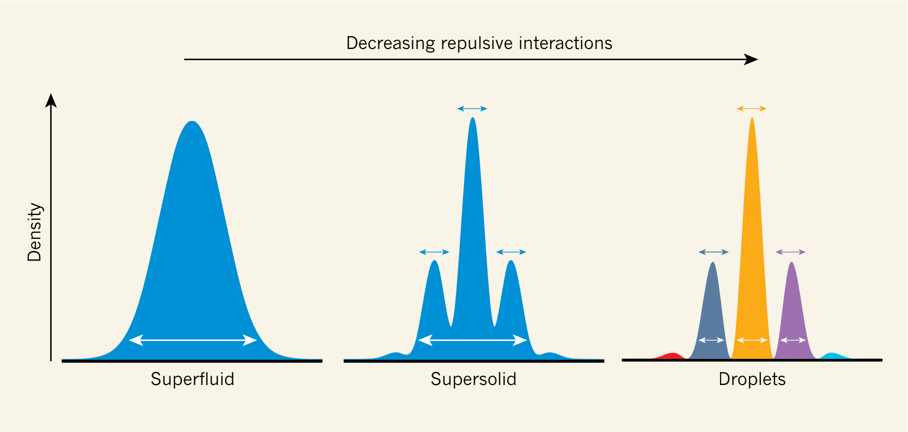


**Figure 1 | Density distributions in a superfluid, a supersolid and isolated droplets.** Three teams[6–8] report experiments on ultracold gases of magnetically dipolar atoms that exist as a superfluid (a fluid that flows without friction). As repulsive interactions between these atoms are decreased, the superfluid becomes a supersolid — a state of matter in which superfluidity and a crystal structure coexist. As the interactions are further decreased, the system forms a collection of isolated droplets (shown by the differently coloured density peaks). The white arrows indicate the regions over which the fluid can flow; the coloured arrows indicate how the crystal structure can move. In the supersolid, there is flow across the entire fluid that is independent of the crystal-structure motion.

In experiments on dipolar BECs, the energy of this roton can be tuned by using an external magnetic field to adjust repulsive short-range interactions.

Much like the cat in Schrödinger's classic thought experiment, a BEC is a quantum object that can exist in two different quantum states simultaneously — a superposition. The existence of a roton allows a BEC to more easily occupy a superposition of two different momenta, effectively moving left and right at the same time. On average, the BEC is stationary, but owing to the wave nature of quantum mechanics, the left-moving and right-moving parts of the system interfere. This interference generates a diffraction pattern, resulting in a periodic arrangement of atoms. In a supersolid, this superposition is the lowest-energy configuration.

Previous reports of supersolids or supersolid-like states in BECs used external influences to produce such a superposition[3–5,10]. What sets dipolar BECs apart from other cold-atom experiments is that no external influence is needed to generate the roton. The emergent crystalline structure spontaneously breaks translational symmetry — the symmetry that is associated with the system being uniformly smooth — in such a way that the crystal structure is free to move and vibrate. This spontaneous symmetry breaking is associated with the emergence of excitations called Higgs and Goldstone modes, which are of fundamental importance in both condensed-matter and high-energy physics.

Sound at low temperature in these exotic systems is characterized by such symmetry breaking, which underpins much of modern physics. In the early 1960s, it was shown that when a system spontaneously breaks a fundamental symmetry, such as that of translation, long-lived, low-energy excitations (sound modes) emerge[11,12]. In the standard model of particle physics, symmetry breaking has a key role in the emergence of light particles such as pions, which are responsible for nuclear interactions, and in the Higgs mechanism, which is responsible for much of the mass in the Universe.

What makes supersolids interesting is that two symmetries are simultaneously broken, resulting in two Goldstone modes. The nature of these two modes can be understood separately: normal sound in the superfluid is associated with the superfluid flow of the BEC, whereas the supersolid sound mode is associated with oscillation of the crystal structure. In practice, the use of an external trap causes these modes to be coupled and discretized. A supersolid has the necessary character of superfluid flow across the entire fluid, independent of the crystal-structure oscillation (Fig. 1). The main goals of the three current papers were to directly observe the Goldstone mode associated with supersolid formation and to distinguish it from the mode related to superfluidity.

> ***"The experiments show what happens when we shake a supersolid, but what happens when we stir or spin it?"***

Guo and colleagues study the first discrete excitation of the sound modes — the sloshing mode of the supersolid in the authors' bowl-like trap. Unfortunately, this mode has a low excitation energy, and therefore moves so slowly that observing it directly would take longer than the lifetime of the supersolid. However, the reported correlation between superfluid displacement and crystal displacement sampled over many iterations indicates that if the superfluid sloshes one way, the crystal tends to move the other way. This result provides convincing evidence for simultaneous superfluidity and crystalline structure, as contrasted with the case in which the system forms independent droplets and the correlation is absent.

Tanzi *et al.* and Natale *et al.* observe a different discrete Goldstone mode, known as a breathing mode. Like an accordion, a supersolid breathing mode is one in which the superfluid and the crystal compress and decompress, but at different frequencies. The authors extract these two oscillations by monitoring the spacing and relative magnitudes of the density peaks as the system is pushed from the regular superfluid regime into the supersolid one. They show that the two oscillation frequencies grow more disparate as the independent-droplet regime is approached.

These three studies are a major step forward as experiments start to probe the properties of supersolids. At the current stage, the restricted size of the observed density modulation (consisting of about three or four linked droplets) and the limited lifetime of the dipolar supersolids pose challenges. However, experimental efforts are already under way to circumvent these issues. In the future, the study of vorticity (how a superfluid forms tornado-like structures) will shed light on the fluid properties of supersolids. The current experiments show what happens when we shake a supersolid, but what happens when we stir or spin it?

Supersolids are also likely to play a key part in our understanding of pulsars (rapidly rotating stellar remnants called neutron stars), making observation in terrestrial experiments even more valuable. Although hot by human standards, neutron stars are cold on nuclear-physics scales, and are expected to contain several forms of superfluid, from neutron superfluids in their crust to 'colour superconductors' in their core. An exotic type of supersolid mechanism[13] predicted in the 1960s might be needed to explain puzzling observations in pulsars. With this emerging generation of experiments, there is now a solid future for the study of supersolids. ■

**Sean M. Mossman** *is in the Department of Physics and Astronomy, Washington State University, Pullman, Washington 99164, USA.*
*e-mail: sean.mossman@wsu.edu*

1. Andreev, A. F. & Lifshitz, I. M. *Sov. Phys. JETP* **29**, 1107–1113 (1969).
2. Kim, D. Y. & Chan, M. H. W. *Phys. Rev. Lett.* **109**, 155301 (2012).
3. Léonard, J., Morales, A., Zupancic, P., Esslinger, T. & Donner, T. *Nature* **543**, 87–90 (2017).
4. Léonard, J., Morales, A., Zupancic, P., Donner, T. & Esslinger, T. *Science* **358**, 1415–1418 (2017).
5. Li, J.-R. *et al. Nature* **543**, 91–94 (2017).
6. Tanzi, L. *et al. Nature* **574**, 382–385 (2019).
7. Guo, M. *et al. Nature* **574**,386–389 (2019).
8. Natale, G. *et al. Phys. Rev. Lett.* **123**, 050402 (2019).
9. Santos, L., Shlyapnikov, G. V. & Lewenstein, M. *Phys. Rev. Lett.* **90**, 250403 (2003).
10. Bersano, T. M. *et al. Phys. Rev. A* **99**, 051602(R) (2019).
11. Goldstone, J. *Nuovo Cimento* **19**, 154–164 (1961).
12. Nambu, Y. *Phys. Rev.* **117**, 648–663 (1960).
13. Kinnunen, J. J., Baarsma, J. E., Martikainen, J.-P. & Törmä, P. *Rep. Prog. Phys.* **81**, 046401 (2018).

MATTERS ARISING

https://doi.org/10.1038/s41586-019-1623-3

# Reversion after replacement of mitochondrial DNA

Gavin Hudson[1]*, Yuko Takeda[1] & Mary Herbert[1,2]*

**ARISING FROM:** E. Kang et al. *Nature* https://doi.org/10.1038/nature20592 (2016).

The risk of transmitting deleterious mutations in mitochondrial DNA (mtDNA) from mother to child may be reduced by mitochondrial replacement, which involves transplanting the nuclear DNA from the egg of a woman containing such a mutation into an enucleated egg from an unaffected donor[1]. Although embryonic stem cell lines derived from mitochondrial replacement embryos can revert to the mitochondrial genome of the nuclear donor[1], a study that involved eggs from women who carry pathogenic mutations reported that mutant mtDNA was efficiently eliminated and did not re-emerge in embryonic stem cell lines[2]. Contrary to this, we find that two embryonic stem cell lines derived by Kang et al.[2] exhibited reversion to a pathogenic variant, which was present in the mitochondrial genome of a nuclear donor assigned to the control group. Although the clinical relevance is unclear, reversion in embryonic stem cell lines remains a consideration for therapeutic applications of mitochondrial replacement, and accurate reporting of the research that underpins these applications is therefore essential.

The mitochondrial replacement procedure involves transplanting the nuclear genome in a karyoplast containing a small amount of cytoplasm. The small amount of mtDNA contained in the karyoplast generally accounts for less than 2% of the mtDNA content of mitochondrial replacement embryos[2–4]. Despite this, around 15% of embryonic stem (ES) cell lines derived from such embryos show complete reversion to the mitochondrial genome of the nuclear donor during proliferation in vitro[2–4]. Because the majority of mitochondrial replacement studies are based on mtDNA sequence variants that are not pathogenic, evidence for the fate of pathogenic variants in ES cell lines derived from mitochondrial replacement embryos is limited. The study published by Kang et al.[2] was therefore of considerable interest. These authors reported that three of the women who donated eggs for their study were heteroplasmic for pathogenic mutations in their mtDNA. The study also included eggs from 11 women who were screened to confirm the absence of mtDNA mutations. Mitochondrial replacement was performed by transplanting the metaphase II spindle from eggs that carried pathogenic mtDNA mutations into enucleated eggs from unaffected donors, as well as between eggs that were donated by women who were deemed to be free of pathogenic mutations.

Consistent with others[3,4], Kang et al.[2] found that reversion to the mitochondrial genome of the nuclear donor occurred in a minority of ES cell lines ($n = 3$ out of 18). This included two lines (ST-ES7 and ST-ES8) that were reported to originate from mitochondrial replacement procedures involving eggs donated by women who did not carry inherited pathogenic mtDNA mutations. The remaining reverted ES cell line (3243ST-ES1) was derived from a mitochondrial replacement procedure in which the nuclear donor was a carrier of the m.3243A>G mtDNA mutation. However, the authors reported that reversion in this ES cell line involved the wild type, rather than the mutated mtDNA. They therefore concluded that mitochondrial genomes that carry pathogenic mutations were efficiently eliminated by spindle transfer[2].

Analysis of the mtDNA sequence data published in Kang et al.[2] indicates that one of the egg donors (ED5, of haplogroup U5a) included in the control group is homoplasmic for a pathogenic mtDNA mutation. The mutation (m.14484T>C) is in the *MT-ND6* gene (Fig. 1) and causes Leber's hereditary optic neuropathy (LHON)[5,6]. The authors derived three ES cell lines from mitochondrial replacement embryos, in which the eggs of this woman were used as nuclear donors. Crucially, these cell lines included two of the ES lines (ST-ES7 and ST-ES8) in which reversion to the mitochondrial genome of the nuclear donor was observed (Fig. 1b). The authors reported that by passage 2–3 the karyoplast mitochondrial genome accounted for 81% and 94% of the total mtDNA in ST-ES7 and ST-ES8 respectively, and that both lines were homoplasmic for the U5a haplogroup by passage 9–10 (ref. [2]). Thus, both ES lines became homoplasmic for the mitochondrial genome that carries the m.14484T>C mutation during proliferation in vitro.

The m.14484T>C mutation is one of three common primary mutations that are known to cause LHON[7]. LHON was first described in 1871[8] and, just over a century later, became the first disease that was proven to be caused by a point mutation in mtDNA[9]. The phenotypic hallmark of LHON is a degeneration of retinal ganglion cells that results in an acute loss of central vision[7]. Unlike most pathogenic mutations of mtDNA, the mutations that cause LHON are typically homoplasmic[7,10]. However, penetrance of the disease is variable. It shows a strong gender bias; there is an approximately 50% chance of blindness in males, compared with about 10% in females[7]. In addition, penetrance is thought to be modulated by factors such as environmental stressors and background genetic effects, including mtDNA haplogroup[10,11]. Globally, the m.14484T>C mutation is the second-most common cause of LHON[12]; however, there is considerable variation in the frequency worldwide. For example, there is an unusually high frequency of the m.14484T>C mutation in cases of LHON among populations of French Canadian descent, owing to a founder effect that dates from the 17th century[11]. Kang et al.[13] note in their corrections to their original paper, and in the accompanying Reply, that the carrier of LHON was assigned to the control group because she had no symptoms or family history of this disease. However, clinical guidelines clearly indicate that the absence of a family history of blindness does not preclude a diagnosis of LHON[7].

Kang et al.[2] also investigated the potential drivers of reversion in ES cell lines. They identified a candidate polymorphism in the mtDNA conserved sequence box II (CSBII), which contains sequences that are involved in generating mtDNA replication primers[14]. In the case of the reverted ES cell lines ST-ES7 and ST-ES8, the authors proposed that the CSBII G6AG8 variant in the mtDNA of the nuclear donor conferred a replicative advantage, relative to the G5AG8 variant that was present in the mtDNA of the mitochondrial donor. They therefore proposed

---

[1]Wellcome Centre for Mitochondrial Research, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. [2]Newcastle Fertility Centre, Newcastle upon Tyne, UK.
*e-mail: gavin.hudson@newcastle.ac.uk; mary.herbert@newcastle.ac.uk

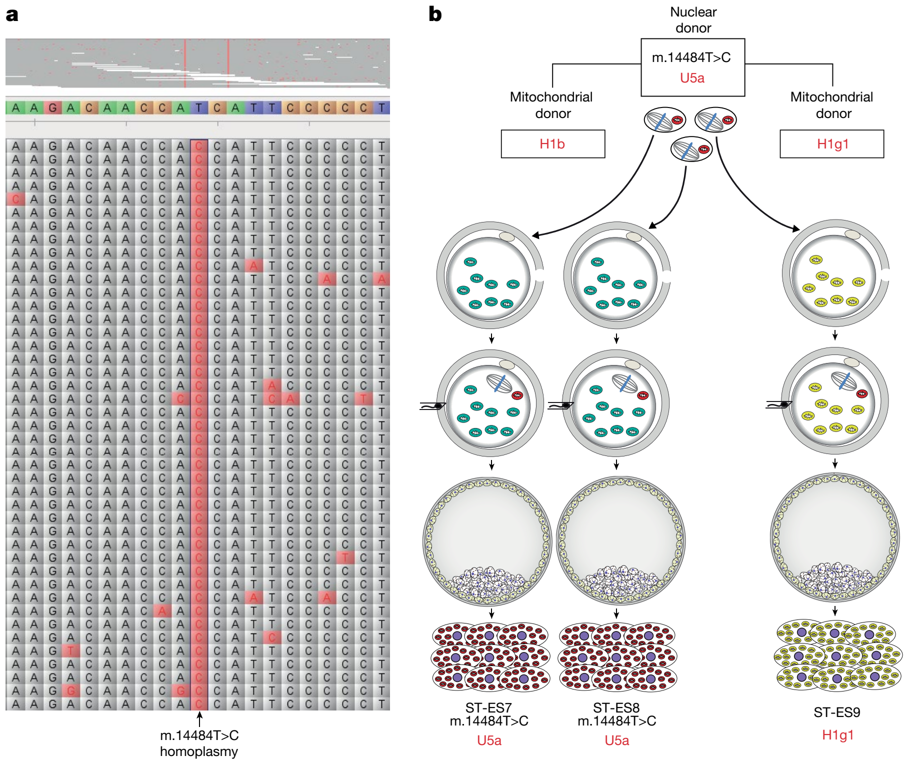


**Fig. 1 | Reversion to a pathogenic mutation in mtDNA, in ES cell lines derived from mitochondrial replacement embryos.** **a**, Next-generation sequencing reads, covering m.14484T>C. The coverage density is 9,859× (top), which corresponds to the revised Cambridge Reference Sequence base pair (m.14484T), and a proportional down-sampling of reads that cover the primary LHON-causing mutation m.14484T>C (bottom, in which C is 9,626× or 99.7% and T is 3× or 0.33%). **b**, Schematic showing egg donor combinations that were used in the mitochondrial replacement procedures that gave rise to three ES cell lines in which the mitochondrial genome of the nuclear donor contained the m.14484T>C variant. mtDNA haplogroups are shown in red. Two ES cell lines (ST-ES7 and ST-ES8) reverted to the haplogroup of the nuclear donor, and become homoplasmic for m.14484T>C.

that the risk of reversion could be reduced by matching donor–recipient pairs according to the similarity of their CSBII haplotypes. Our analysis of the mtDNA sequence data in Kang et al.[2] indicates that the donor–recipient combinations of CSBII variants were identical between ST-ES7, ST-ES8 and ST-ES9. However, in contrast to ST-ES7 and ST-ES8, the ST-ES9 cell line did not revert (Fig. 2a). To further investigate the relevance of matching CSBII haplotypes, we analysed data from published ES cell lines that were derived from mitochondrial replacement embryos[2–4] ($n = 31$, listed in Extended Data Table 1; see Supplementary Methods for further details). Our findings indicate that the incidence of reversion is similar between matched and unmatched CSBII haplotypes (Fig. 2b, c). We therefore conclude that there is insufficient evidence to support the proposal that matching donor–recipient pairs on the basis of similarity of the CSBII haplotype

**a**

| ES cell line | CSBII haplotypes | Result |
|---|---|---|
| ST-ES7 | G6AG8:G5AG8 | Reverted |
| ST-ES8 | G6AG8:G5AG8 | Reverted |
| ST-ES9 | G6AG8:G5AG8 | Non-reverted |

**b**

CSBII haplotype of mitochondrial donor

| CSBII haplotype of nuclear donor | G5AG7 | G5AG8 | G6AG7 | G6AG8 | G6AG9 |
|---|---|---|---|---|---|
| G5AG7 | 3:1 | – | – | – | – |
| G5AG8 | – | – | – | 3:0 | – |
| G6AG7 | 2:0 | – | 3:1 | 6:0 | – |
| G6AG8 | – | 1:2 | 3:1 | 4:0 | – |
| G6AG9 | 1:0 | – | – | – | – |

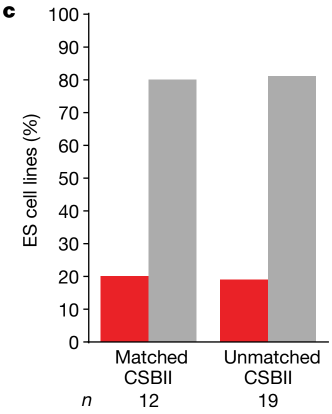


**Fig. 2 | Effect of CSBII haplotypes on reversion in ES cell lines after mitochondrial replacement.** **a**, CSBII polymorphisms in ES cell lines reported in Kang et al.[2], according to whether or not they reverted to the mitochondrial genome of the nuclear donor. **b**, Table showing the ratio of non-reverted to reverted (black:red) ES cell lines from all published lines derived from mitochondrial replacement embryos[2–4] **c**, The bar chart shows the proportion of ES cell lines derived from mitochondrial replacement embryos that reverted to the mitochondrial genome of the nuclear donor (red bars) or not (grey bars), according to whether the nuclear and mitochondrial donors had matched ($n = 12$) or unmatched ($n = 19$) CSBII polymorphic variants. There was no significant difference in the proportions (Fisher's exact test, $P > 0.05$).

reduces the risk of reversion. In support of this, we note that—in the accompanying Reply—Kang et al. now propose an alternative strategy that involves matching donor–recipient pairs on the basis of differences in the number of guanosine residues in the CSBII sequence.

In summary, a donor whose eggs were used for control experiments in Kang et al.[2] is homoplasmic for the m.14484T>C mtDNA variant, which causes LHON. Transplantation of nuclear DNA from the eggs of this donor resulted in mitochondrial replacement embryos that gave rise to three ES cell lines, two of which became homoplasmic for the m.14484T>C mutation. Because mutations that cause LHON are typically homoplasmic, mitochondrial replacement is currently the only means of reducing the risk of transmission for women with such a mutation who wish to have their own genetically related child. Although it is unknown whether reversion occurs during development in vivo, this possibility remains a consideration for the clinical translation of mitochondrial replacement. Taken together with previous reports[3,4], the finding that ES cell lines derived from mitochondrial replacement embryos reverted to a mitochondrial genome that carries a pathogenic variant underscores the importance of minimizing the contribution of the mtDNA of the nuclear donor after mitochondrial replacement. Finally, our analysis (together with the accompanying Reply) does not support the previous recommendation[2], which advocated matching donor–recipient pairs according to the similarity of their CSBII haplotype.

## Data availability

Previously unpublished[4] mtDNA sequence data from ES cell lines are available online (https://doi.org/10.5281/zenodo.3349761). Additional previously published data are available from the respective publications[2,3]. Further supporting sequence data from ref. [3] were provided to us by one of the authors (D. Egli).

Received: 13 May 2019; Accepted: 14 August 2019;

1. Greenfield, A. et al. Assisted reproductive technologies to prevent human mitochondrial disease transmission. *Nat. Biotechnol.* **35**, 1059–1068 (2017).
2. Kang, E. et al. Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations. *Nature* **540**, 270–275 (2016).
3. Yamada, M. et al. Genetic drift can compromise mitochondrial replacement by nuclear transfer in human oocytes. *Cell Stem Cell* **18**, 749–754 (2016).
4. Hyslop, L. A. et al. Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* **534**, 383–386 (2016).
5. Mackey, D. A. et al. Primary pathogenic mtDNA mutations in multigeneration pedigrees with Leber hereditary optic neuropathy. *Am. J. Hum. Genet.* **59**, 481–485 (1996).
6. Carelli, V. et al. Biochemical features of mtDNA 14484 (ND6/M64V) point mutation associated with Leber's hereditary optic neuropathy. *Ann. Neurol.* **45**, 320–328 (1999).
7. Yu-Wai-Man, P. & Chinnery, P. F. in *GeneReviews* (eds Adam, M. P. et al.) https://www.ncbi.nlm.nih.gov/books/NBK1174/ (Univ. of Washington, 2000).
8. Leber, T. Ueber hereditäre und congenital-angelegte Sehnervenleiden. *Albr. von Graefes Arch. für Ophthalmol.* **17**, 249–291 (1871).
9. Wallace, D. C. et al. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science* **242**, 1427–1430 (1988).
10. Wallace, D. C. & Lott, M. T. in *Pharmacology of Mitochondria* (eds Singh, H. & Sheu, S.-S.) 339–376 (Springer International, 2017).
11. Hudson, G. et al. Clinical expression of Leber hereditary optic neuropathy is affected by the mitochondrial DNA-haplogroup background. *Am. J. Hum. Genet.* **81**, 228–233 (2007).
12. Meyerson, C., Van Stavern, G. & McClelland, C. Leber hereditary optic neuropathy: current perspectives. *Clin. Ophthalmol.* **9**, 1165–1176 (2015).
13. Kang, E. et al. Author Correction: Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations. *Nature* **567**, E5–E9 (2019).
14. Agaronyan, K., Morozov, Y. I., Anikin, M. & Temiakov, D. Replication-transcription switch in human mitochondria. *Science* **347**, 548–551 (2015).

**Acknowledgements** Work in the laboratory of G.H. and M.H. is funded by Wellcome (203105/Z/16/Z). G.H. is a Newcastle University Research Fellow and receives funding from the Michael J. Fox Foundation (ID: 15643).

**Author contributions** G.H., Y.T. and M.H. performed the analysis. G.H. and M.H. wrote the manuscript with input from Y.T.

**Competing interests** The authors declare no competing interests.

**Additional information**
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1623-3.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
**Correspondence and requests for materials** should be addressed to G.H. or M.H.

**Extended Data Table 1 | mtDNA CSBII haplotypes of nuclear and mitochondrial donors**

| *Mitochondrial Replacement Experiment* | *Reference* | *ES Lines (n)* | *Reverted ES Lines* | *Donor CSBII Haplotype Nuclear : Mitochondrial* |
|---|---|---|---|---|
| ST-ES1 | 2 | 1 | 0% | G6AG8:G6AG8 |
| ST-ES2 & ST-ES3 | 2 | 2 | 0% | G6AG8:G6AG8 |
| ST-ES4 | 2 | 1 | 0% | G6AG7:G5AG7 |
| ST-ES5 & ST-ES6 | 2 | 2 | 0% | G5AG8:G6AG8 |
| ST-ES7 & ST-ES8 | 2 | 2 | 100% | G6AG8:G5AG8 |
| ST-ES9 | 2 | 1 | 0% | G6AG8:G5AG8 |
| ST-ES10 | 2 | 1 | 0% | G5AG8:G6AG8 |
| ST-ES11-13 | 2 | 3 | 0% | G5AG7:G5AG7 |
| ST-ES14 | 2 | 1 | 0% | G6AG7:G6AG8 |
| ST-ES15 | 2 | 1 | 0% | G6AG8:G6AG7 |
| NA | 2 | 0 | - | Data Unavailable |
| 13513ST-ES | 2 | 1 | 0% | G6AG7:G6AG7 |
| 3243ST-ES1 & 3242ST-ES2 | 2 | 2 | 50% | G6AG7:G6AG7 |
| MR-PS1 | 2 | 0 | - | Data Unavailable |
| MR-PS2 | 2 | 0 | - | Data Unavailable |
| MR-PS3 | 2 | 0 | - | Data Unavailable |
| MR-PS4 | 2 | 0 | - | Data Unavailable |
| MR-PS5 | 3 | 1 | 0% | G6AG9:G6AG7 |
| MR-PS6 & MR-PS7 | 3 | 2 | 0% | G6AG8:G6AG7 |
| MR-PS8 | 3 | 1 | 0% | G6AG7:G6AG7 |
| MR-PS9 & MR-PS10 | 3 | 2 | 0% | G6AG7:G6AG8 |
| MR-PS11 | 3 | 1 | 0% | G6AG7:G6AG8 |
| MR-PS12 | 3 | 1 | 100% | G6AG8:G6AG7 |
| 31PNT | 4 | 1 | 0% | G6AG8:G6AG8 |
| 36PNT | 4 | 1 | 100% | G5AG7:G5AG7 |
| 45PNT | 4 | 1 | 0% | G6AG7:G5AG7 |
| 47PNT | 4 | 1 | 0% | G6AG7:G6AG8 |
| 55PNT | 4 | 1 | 0% | G6AG7:G6AG8 |

These haplotypes are taken from previously published[2–4] ES cell lines derived from mitochondrial replacement embryos. In the case of ES cell lines published in Kang et al.[2], the sequence data were downloaded from the Sequence Read Archive at the NCBI (accession PRJNA349068). mtDNA sequence data from the ES cell lines reported in ref. [3] were provided by the authors (D. Egli, personal communication), and sequencing data for ES cell lines reported in ref. [4] were obtained from our own next-generation sequencing data (Y.T. et al., manuscript in review).

MATTERS ARISING

https://doi.org/10.1038/s41586-019-1624-2

# Reply to: Reversion after replacement of mitochondrial DNA

Eunju Kang[1], Amy Koski[2], Paula Amato[2,3], Dmitry Temiakov[4] & Shoukhrat Mitalipov[2,3]*

**REPLYING TO** G. Hudson et al. *Nature* https://doi.org/10.1038/s41586-019-1623-3 (2019)

In the accompanying Comment[1], Hudson et al. argue that one of our research volunteers (ED5, with U5a haplotype), who was assigned to the unaffected-oocyte cohort of donors in our original study[2], should be allocated to the mitochondrial DNA (mtDNA)-disease group because she carries a homoplasmic m.14484T>C variant of mtDNA that is implicated in Leber's hereditary optic neuropathy (LHON). However, for the purposes of our in vitro study, oocyte donors were assigned to the healthy cohort on the basis of clinical criteria alone—that is, the absence of a diagnosis or symptoms of disease caused by mutations in mtDNA. In addition, the inclusion and exclusion criteria of our approved study protocol dictated that, to be considered as possessing a pathogenic mutation in their mtDNA, the research volunteer had to be diagnosed with a maternally transmitted mitochondrial disease. Because this oocyte donor was healthy, and lacked any evidence of LHON in her family for four generations, she was assigned to the healthy-oocyte cohort of donors. However, carriers of 14484T>C variants of mtDNA should not be used as oocyte donors in future clinical applications of mitochondrial replacement therapy—whether the carrier is asymptomatic, or not.

Hudson et al.[1] also dispute our conclusion that sequence polymorphism within the conserved sequence box II (CSBII) region of human mtDNA affecting the efficiency of mtDNA replication may contribute to reversal after mitochondrial replacement therapy. Results from our previous study[2] (which are illustrated in figure 2 in ref. [1]) indicate that, among the three embryonic stem (ES) cell lines with specific maternal–donor CSBII combinations (G6AG8–G5AG8), two lines (66%) reversed back to the maternal haplotype. However, in our original study[2], we generated three additional ES cell lines (ST-ES5, ST-ES6 and ST-ES10) with opposite maternal–donor combinations (that is, G5AG8–G6AG8) but none of these cell lines reversed. To define possible mechanisms that might underlie such outcomes, we measured synthesis of the replication primer by mitochondrial RNA polymerase and found that the

**Table 1 | Effect of maternal and donor CSBII sequences on the reversal of ES cell lines generated by mitochondrial replacement therapy**

| Number of G residues in maternal > donor | | Number of G residues in maternal = donor | | Number of G residues in maternal < donor | |
|---|---|---|---|---|---|
| ES cell lines | CSBII (maternal:donor) | ES cell lines | CSBII (maternal:donor) | ES cell lines | CSBII (maternal:donor) |
| ST-ES7* | G6AG8:G5AG8 | ST-ES1 | G6AG8:G6AG8 | ST-ES4 | G6AG7:G6AG8 |
| ST-ES8* | G6AG8:G5AG8 | ST-ES2 | G6AG8:G6AG8 | ST-ES5 | G5AG8:G6AG8 |
| ST-ES9 | G6AG8:G5AG8 | ST-ES3 | G6AG8:G6AG8 | ST-ES6 | G5AG8:G6AG8 |
| ST-ES15 | G6AG8:G6AG7 | 3243ST-ES1* | G6AG8:G6AG8 | ST-ES10 | G5AG8:G6AG8 |
| | | 3243ST-ES2 | G6AG8:G6AG8 | ST-ES11 | G5AG6:G6AG7 |
| | | NT-ES1 | G6AG8:G6AG8 | ST-ES12 | G5AG6:G6AG7 |
| | | NT-ES2 | G6AG8:G6AG8 | ST-ES13 | G5AG6:G6AG7 |
| | | NT-ES3 | G6AG8:G6AG8 | ST-ES14 | G6AG7:G6AG8 |
| | | NT-ES4 | G6AG8:G6AG8 | 13513ST-ES | G6AG7:G6AG9 |
| | | NT-ES5 | G6AG7:G6AG7 | NT-ES6 | G5AG7:G5AG8 |
| | | NT-ES8* | G6AG7:G6AG7 | NT-ES7 | G6AG7:G6AG8 |
| MR-PS12* | G6AG8:G6AG7 | MR-PS1 | G5AG7:G5AG7 | MR-PS11 | G6AG7:G6AG8 |
| | | MR-PS2 | G5AG7:G5AG7 | | |
| | | MR-PS3 | G5AG7:G5AG7 | | |
| | | NT5 | G6AG7:G6AG7 | | |
| | | NT6* | G6AG7:G6AG7 | | |
| | | NT8* | G6AG7:G6AG7 | | |
| 3/5 (60%) | | 4/17 (24%) | | 0/12 (0%) | |

The cell lines MR-PS1, MR-PS2, MR-PS3, MR-PS11, MR-PS12, NT5, NT6 and NT8 are from ref. [3]; the remaining 26 cell lines are from ref. [2]. CSBII sequence information for these 34 cell lines has been deposited in public databases, and was used to generate this table. The number of guanosine residues was counted in the first (m.311–315) and second (m.303–309) set of G repeats within the CSBII region. The reversal rate in the maternal > donor group is significantly higher than in the maternal < donor group (Fisher's exact test, $P < 0.05$).

*Reversed ES cell lines.

---

[1]Department of Convergence Medicine and Stem Cell Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. [2]Center for Embryonic Cell and Gene Therapy, Oregon Health & Science University, Portland, OR, USA. [3]Division of Reproductive Endocrinology, Department of Obstetrics and Gynecology, Oregon Health & Science University, Portland, OR, USA. [4]Department of Biochemistry and Molecular Biology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA. *e-mail: mitalipo@ohsu.edu

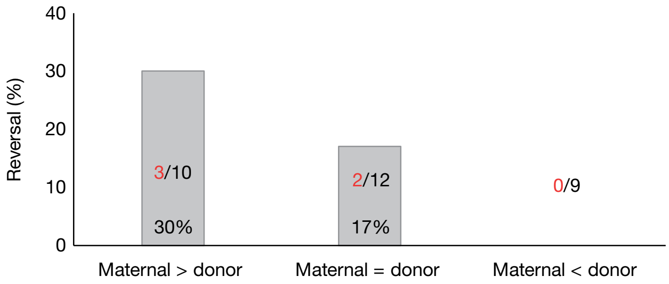


**Fig. 1 | Distribution of reversed ES cell lines based on CSBII sequence.** Distribution of ES cell lines that reversed after mitochondrial replacement therapy (shown in red font), based on number of guanosine residues in the CSBII region of combinations of maternal and donor mtDNA. Cell line information is from figure 2b in ref. [1].

deletion of a single guanosine residue in the CSBII region (G5AG8 versus G6AG8) results in a fourfold reduction of replication primer synthesis (as shown in figure 3b of ref. [2]).

Two other cell lines from our study that reversed after mitochondrial replacement therapy carried identical maternal–donor CSBII sequences (Table 1), which suggests that additional sequence-independent mechanisms affect reversal. Notwithstanding this finding, we proposed that a paradigm of selecting donor mtDNA on the basis of the CSBII sequence could preclude, or minimize, the possibility of reversal and the resurgence of disease caused by mutations in mtDNA in children born after mitochondrial replacement therapy[2].

Extended Data Table 1 in the accompanying Comment[1] analyses the contribution of CSBII combinations to reversal in 31 ES cell lines that have been generated in three recent studies[2–4]. Of note, published sequence information is available for 34 ES cell lines, but from only two of the studies[2,3] (Table 1); the CSBII sequence information from ref. [4] has not been published.

In figure 2c of the accompanying Comment[1], Hudson et al. compare the frequency of reversal between two groups: (1) matched, in which maternal and donor CSBII sequences are identical; and (2) unmatched, in which the maternal and donor CSBII sequences differ. On the basis of this comparison, they conclude that the incidence of reversal is similar between these two groups, and thus that the risk of reversal is unlikely to be reduced by CSBII matching.

However, we and others[2,5,6] have found that a reduction in the number of guanosine residues in CSBII region might result in reduced replication. Therefore, a paradigm for selecting maternal and donor mtDNA on the basis of the CSBII region could reduce incidences of reversal if the total number of guanosine residues within the CSBII region in donor mtDNA is greater than it is in the maternal mtDNA; that is, if G6AG8 > G5AG8 (14 > 13), G6AG7 > G5AG7 (13 > 12), G6AG8 > G6AG7, G6AG9 > G5AG7 and so on (Table 1). Therefore, it will be more informative to distribute the samples from figure 2b in the accompanying Comment[1] into three groups on the basis of the ratio of guanosine residues in the CSBII region in maternal and donor mtDNA: (1) maternal > donor, (2) maternal = donor and (3) maternal < donor. The analysis we perform here (Fig. 1) demonstrates that if the CSBII region of maternal mtDNA contains more guanosine residues than it does in donor mtDNA, 3 out of 10 cell lines reversed (30%). In the second group, in which the maternal:donor ratio was equal, 2 out of 12 cell lines reversed (17%). However, if there were fewer guanosine residues in the CSBII region of maternal mtDNA than in same region of the donor mtDNA, 0 out of 9 cell lines reversed (Fig. 1).

Although the number of combinations that have thus far been tested is small and requires expansion, our analysis of published studies supports the conclusion that the selection of donor mtDNA on the basis of the CSBII sequence (that is, with fewer guanosine residues in the CSBII of the maternal mtDNA as compared to donor mtDNA) could be beneficial to avoid reversal. This is a precautionary issue, given safety considerations in ongoing mitochondrial replacement therapy trials at Newcastle University (UK) and other clinics around the world[7]. We find the arguments advanced in the accompanying Comment[1]—that mtDNA matching may restrict the availability of egg donors for mitochondrial replacement therapy—inadequate, given the possibility of disease recurrence in children after mitochondrial replacement therapy. The selection of CSBII haplotype in donors suggested here should not be so onerous as to impede the progress of the trials.

We realize that differences in the sequence of the CSBII region are not the only factors that drive reversal and, in our original study, suggested that other regulatory sequences in the control region of mtDNA "may, subsequently, confer replicative advantage" and "contribute to replication bias of a particular mtDNA haplotype"[2]. We also addressed in our recent review[8] that other sequence-independent mechanisms are likely to affect reversal, and proposed epigenetic mechanisms that might be involved.

The authors E.K., A.K., P.A., D.T. and S.M. are solely responsible for this Reply; other authors from the original Letter did not contribute to this response and are not listed here.

## Data availability

All data presented in this study have previously been published[2–4].

1. Hudson, G., Takeda, Y. & Herbert, M. Reversion after replacement of mitochondrial DNA. *Nature* https://doi.org/10.1038/s41586-019-1623-3 (2019).
2. Kang, E. et al. Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial DNA mutations. *Nature* **540**, 270–275 (2016).
3. Yamada, M. et al. Genetic drift can compromise mitochondrial replacement by nuclear transfer in human oocytes. *Cell Stem Cell* **18**, 749–754 (2016).
4. Hyslop, L. A. et al. Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* **534**, 383–386 (2016).
5. Tan, B. G., Wellesley, F. C., Savery, N. J. & Szczelkun, M. D. Length heterogeneity at conserved sequence block 2 in human mitochondrial DNA acts as a rheostat for RNA polymerase POLRMT activity. *Nucleic Acids Res.* **44**, 7817–7829 (2016).
6. Agaronyan, K., Morozov, Y., Anikin, M. & Temiakov, D. Replication-transcription switch in human mitochondria. *Science* **347**, 548–551 (2015).
7. Thomson, H. First 3-parent baby born in clinical trial to treat infertility. https://www.newscientist.com/article/2199441-first-3-parent-baby-born-in-clinical-trial-to-treat-infertility/ (2019).
8. Wolf, D. P., Hayama, T. & Mitalipov, S. Mitochondrial genome inheritance and replacement in the human germline. *EMBO J.* **36**, 2659 (2017).

**Author contributions** E.K., A. K. and S.M. drafted the first version of the manuscript. E.K., A. K., P.A., D.T., and S.M. provided feedback and edited the final version.

**Competing interests** S.M. is a co-founder and shareholder of Mitogenome Therapeutics, Inc. The other authors declare no competing financial interests

**Additional information**

**Correspondence and requests for materials** should be addressed to S.M.

REVIEW

https://doi.org/10.1038/s41586-019-1643-z

# Subducting carbon

Terry Plank[1]* & Craig E. Manning[2]

**A hidden carbon cycle exists inside Earth. Every year, megatons of carbon disappear into subduction zones, affecting atmospheric carbon dioxide and oxygen over Earth's history. Here we discuss the processes that move carbon towards subduction zones and transform it into fluids, magmas, volcanic gases and diamonds. The carbon dioxide emitted from arc volcanoes is largely recycled from subducted microfossils, organic remains and carbonate precipitates. The type of carbon input and the efficiency with which carbon is remobilized in the subduction zone vary greatly around the globe, with every convergent margin providing a natural laboratory for tracing subducting carbon.**

In addition to its familiar cycling between the terrestrial biosphere and atmosphere, carbon moves from microfossils on the seafloor to erupting volcanoes and deep diamonds, in a cycle driven by plate tectonics. Subduction links surface biological processes with the deep Earth, creating a planet suffused with the signature of life. The fate and flux of carbon vary from trench to trench, as every subducting slab delivers to the mantle a singular mix of organic and inorganic carbon that heats and pressurizes at a particular rate determined by its age and convergence speed. The inherent reactivity of carbon in each input package and the pressure–temperature–time path experienced by each slab drive chemical reactions that separate carbon into two parts: one that is carried deeper into Earth and one that returns to the surface. Here we discuss (I) the geological happenstance that delivers carbon to deep sea trenches today, (II) the reactions and movements in the subduction zone that mobilize carbon into fluids and melts, (III) the carbon that continues deep into the mantle to form molten carbonates and diamonds, and (IV) the carbon that escapes such a deep descent to be eventually emitted from arc volcanoes (Fig. 1).

Although excellent reviews exist on the global consequence of the deep carbon cycle over Earth's history[1–6], our focus here is on the variability in current subduction systems, where carbon inputs and outputs can reveal the physicochemical processes occurring in the subduction zone that ultimately drive the directionality of the cycle. We highlight end-member subduction zones with distinct inputs or pathways, such as Tonga (little sedimentary carbon) and Cascadia (hot slab), each presenting a different natural recycling experiment worthy of focused study. We also show how recent advances in both laboratory and computational approaches now lead to predictions of how and when carbon mobilizes into upward-moving fluids or remains in solid form in the downgoing plate. Finally, we present recent evidence from satellites, portable instruments and melt inclusions that suggests that most of the carbon in arc magmas is recycled from the subduction zone.

## The global carbon flux perspective

Earth exhales carbon at volcanoes. Current estimates place the total volcanic outgassing rate at 79 ± 9 (uncertainty 1$\sigma$) megatons of carbon per year (Mt C yr$^{-1}$), with similar amounts emitted at mid-ocean ridges and subduction zones, and the greatest quantities emitted diffusely in intraplate volcanic regions (Box 1). Uncertainties in estimates of volcanic emission rates are quickly declining owing to long-term satellite observation[7] and direct sampling of volcanic plumes and diffuse emanations[8–10]. The spatial variability in volcanic carbon is enormous, however, and currently six volcanoes are estimated to control the global budget of direct emissions (Nyiragongo, Popocateptl, Etna, Ambrym, Bagana and Aoba; together contributing more than 5 Mt C yr$^{-1}$)[11]. The largest volcanic emissions, however, may be emanating diffusely from calderas and faults[10]. Although current anthropogenic fluxes (about 9,500 Mt C yr$^{-1}$)[12] dwarf volcanic ones, the long-term fluxes of carbon to the atmosphere have been dominated by volcanic sources over most of Earth's history. The rate of supply of mantle carbon is governed by tectonic rates of plate spreading and convergence, as well as the plume flux from the deeper mantle, punctuated by large igneous provinces that supply bursts of $CO_2$ in short periods of time.

Deep-Earth carbon fluxes are not a one-way street, however. There is an equally substantial, highly uncertain and spatially variable carbon flux that enters the mantle with oceanic plates at subducting zones. This carbon is derived largely from the ocean, in the form of carbonate shells and remains of marine organisms, as well as carbonate in the oceanic lithosphere. Terrestrial organic carbon is also washed onto the seafloor by large rivers. The fate of subducting carbon has a profound effect on Earth's evolution, and depends on the efficiency with which carbon is returned to Earth's surface by devolatilization reactions in the subducting slab and by upward transport in magmas that supply volcanic arcs. We will refer to this return flux of carbon to Earth's surface as 'recycling'. A recycling efficiency of 0% means that none of the carbon that is subducted returns to Earth's surface, but instead is sequestered in the mantle for geologically significant durations.

A low recycling efficiency could have planetary consequences. For example, Earth's surface has a distinctively high H/C ratio, about an order of magnitude higher than Earth's chondritic building blocks[13]. Is this due to more efficient subduction recycling of $H_2O$ than that of C over Earth's history[4]? The deep subduction of reduced organic carbon amounts to a loss of electrons and therefore an increase in the oxidative potential of Earth's surface[1]. Has a low recycling efficiency of organic carbon contributed to Earth's oxygen-rich surface[14]? Or is the oxidation of the mantle the net effect of subduction[15], given the higher input flux of carbonate than that of organic carbon? What is the separate fate of these oxidized and reduced forms of carbon?

A high recycling efficiency creates a direct connection between subducting carbon and $CO_2$ supply to the atmosphere. For example, there may have been greater subduction of carbonate at certain times in Earth's history, such as the Mesozoic era, when the seafloor of the shallow Tethys ocean subducted. Did this lead to higher volcanic $CO_2$

---

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA. [2]University of California, Los Angeles, CA, USA. *e-mail: tplank@ldeo.columbia.edu

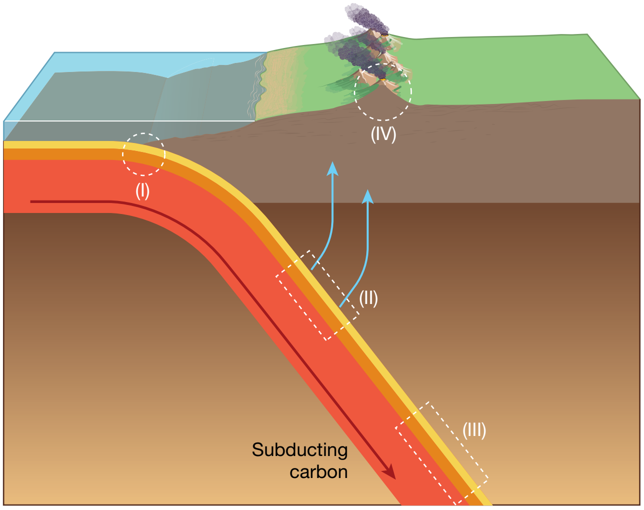


**Fig. 1 | The deep carbon cycle.** (I) Carbon input to subduction zones from the sedimentary (yellow), oceanic crust (orange) and mantle (red) layers in the downgoing plate (dark red arrow). (II) Decarbonation reactions in the subduction zone release carbon to upward-moving fluids and melts (blue arrows). (III) Deeply subducted carbon, potentially forming diamonds. (IV) Emission of recycled carbon to the surface via arc volcanism.

emissions and therefore greater warming[16,17]? Does the climate system respond to $CO_2$ precipitation in the oceanic crust as a sink or source of carbon, and how is this process modulated by plate spreading rates, ocean bottom temperatures and subduction delivery[18,19]? What are the long-term climate consequences of the fact that subduction of carbonate is a relatively recent development, tied to evolution of calcifying marine microorganisms[20]? Participation of carbonate subduction in the deep carbon cycle is arguably a recent phenomenon[5].

It is therefore critical to understand the efficiency of carbon subduction recycling, which is determined by two general approaches. One is to balance the input (subducted) and output (arc volcanic) fluxes of carbon globally[2,3]. The most recent estimates indicate about 30% recycling, although the uncertainties are very large in several fluxes (Box 1). Weak correlations may exist between decarbonation efficiency and slab age[21]. The other approach is to characterize the physicochemical processes in subduction zones that drive recycling, such as fluid and melt generation and transport, coupled with carbon solubility and reaction kinetics, given different carbon feedstocks and pressure–temperature paths. Here we focus on the latter approach and trace carbon through the recycling process, first by considering how and when carbon is deposited on and precipitated in subducting seafloor.

## Subducting carbon

Although the ocean is full of carbon—dissolved as bicarbonate and in the shells and bodies of marine organisms—very little carbon makes it to the deep seafloor to be conveyed to a trench. The carbon that enters subduction zones includes calcium carbonate and reduced organic carbon that exist within the sedimentary, oceanic crust and mantle layers of the incoming plate. Each trench makes unique selections from the carbon menu.

Starting with the lowermost layer, mantle peridotite that forms the bulk of the subducting lithosphere readily hydrates and carbonates if exposed to seawater, forming carbonated serpentinites[22]. However, most peridotite usually resides at least 6 km beneath the seafloor and is not in direct contact with seawater. Faulting and fracturing are necessary to bring mantle rocks to the sea floor or seawater to the mantle. Carbonated serpentinites may form near spreading centres or near trenches. Near spreading centres, extensional faulting is linked to hydration and carbonation reactions, as well as to the precipitation of magnesium and calcium carbonate veins in mantle peridotites[22]. Carbon isotopes in oceanic peridotites reflect mixing between seawater-derived carbonate and reduced carbon[23]. Serpentinization is most pronounced at slow-spreading mid-ocean ridges, but most subducting lithosphere is not formed there, because of the low plate production rate and the preferential existence of subduction zones and fast-spreading crust in the Pacific. Oceanic carbonated serpentinites therefore make a minor contribution to the global carbon input flux to subduction zones[3], although they may be locally important. The volcanism at the South Sandwich margin, where slow-spreading crust is preferentially consumed, is notable for having some of the highest $^{11}B/^{10}B$ ratios among island arcs, a feature that could be derived from the high $^{11}B/^{10}B$ ratio that is typical of seafloor serpentinized peridotites[24]. The South Sandwich margin may thus represent a rare case, in which subducting carbon predominantly resides in oceanic carbonated serpentinite (Box 2).

Near trenches, the subducting plate deforms and fractures because of bending. The resulting faults have been seismically imaged to penetrate into the mantle of the incoming plate, and seismic velocities decrease towards the trench[25], leading to speculation that the extent of hydration due to ingress of seawater could consume an ocean every billion years[26,27]. However, outer-rise serpentinites have never been sampled, and other factors besides hydration—such as fracturing and anisotropy—may explain this reduction in seismic velocity. The extent of carbonation in unsampled outer-rise serpentinites is also unknown, and fluid pathways longer than 5 km may lead to low fluid-to-rock ratios and low carbon transport into the mantle section of the downgoing plate[23]. Indeed, electrical-resistivity imaging does not support extensive reaction of seawater with the mantle section of the incoming plate at the Central America trench[28]. Most global flux estimates for carbonated peridotite are low (Box 1).

In contrast to peridotite carbonation, there is abundant evidence for pervasive carbonation of the near-ridge oceanic crust. Seawater-derived fluids circulate predominantly in the higher-permeability upper-crustal volcanic section, leading to low-temperature precipitation of carbonate minerals. Because these carbonates are largely seawater-derived, they are isotopically heavy ($\delta^{13}C \approx 0‰$, where $\delta^{13}C$ is the deviation of the ratio $^{13}C/^{12}C$ relative to that of Pee Dee belemnite), although some biotic and abiotic processes also lead to $CO_2$ reduction and precipitation of isotopically light carbon ($\delta^{13}C < -20‰$[29,30]; Fig. 2a). Nonetheless, the dominant form of carbon in altered oceanic crust (AOC) is calcium carbonate (calcite and aragonite) that precipitates in veins and vugs, as has been found in samples from a small number (about 15) of drill sites. Carbon uptake occurs near the ridge axis in crust that is 20 Myr old, but surprisingly, AOC older than 80 Myr has higher carbonate content[31]. This may be due to higher bottom-water temperatures in the Cretaceous period promoting greater abiogenic carbonate precipitation[32]. Although low in carbonate, young AOC (less than 10 Myr old) is isotopically light owing to the intense bio-alteration of young crust[30]. Thus, an important prediction for carbon inputs is that old plates (for example, Marianas and Tonga) will have greater AOC carbonate concentrations and higher average $\delta^{13}C$, whereas young plates (for example, Cascadia and Central America) will have little AOC carbonate with lower $\delta^{13}C$ (Fig. 2a).

The sedimentary layer that is deposited on top of the oceanic crust contains dramatically different forms of carbon than the largely inorganic precipitates of the oceanic crust and peridotite. Sediments are the graveyard of marine organisms and the resting place of terrestrial organic remains. Organisms that grow a carbonate shell, such as nannoplankton coccoliths and bottom-dwelling foraminifera, are the richest source of carbon deposits on the seafloor. For example, a 100-m section of nannofossil ooze may contain as much carbon as the entire oceanic crust below it (using average values given in ref. [3]). Marine sediments also contain the organic remains of marine and terrestrial organisms. Although most sediments have less than 1 wt% organic carbon, deep-sea fans can dominate the input flux at some margins. For example, a 1.5-km section of terrigenous turbidites with 0.35 wt% organic C (Fig. 2a) contains more carbon than the average oceanic crust. The balance of marine carbonate ($\delta^{13}C = 0‰$ to $+3‰$) versus organic carbon ($\delta^{13}C = -22‰$ to $-27‰$) has a very large effect on the isotopic composition of subducting carbon (see ref. [33] and Fig. 2a). Carbonate sediments can approach 100% $CaCO_3$ and therefore may contain more than 10 times the carbon of a sediment rich in organic carbon (1% C; Fig. 2a). This is offset by a greater flux of organic carbon-bearing sediments approaching trenches (in thick fans), so the proportion of organic to inorganic carbon subducted globally may be about 20% (ref. [34]).

Although sedimentary carbon has the potential to dominate global input fluxes, it may be entirely absent from some subducting sections. Indeed, the odds are stacked against carbon burial, as most carbonate dissolves and organic carbon oxidizes in the water column before reaching the seafloor. The ocean's cold and corrosive bottom waters are particularly challenging to carbonate survival. The calcite compensation depth (CCD), which marks the transition between carbonate-bearing and carbonate-absent sediments, is about 5,000 m deep in today's oceans. The CCD increases locally if the carbonate flux is high, as occurs in regions of high biological productivity, but it was generally shallower (less than 3,500 m) earlier in the Cenozoic era[35] and the Cretaceous period[36]. Much of the oceanic crust subducting today is old (average age of about 70 Myr)[37] and the combination of a shallower CCD and thermal subsidence with age means that carbonate is rare on the seafloor near trenches. For example, essentially zero sedimentary carbonate is subducted along the Tonga, Central Aleutian and Kuriles–Kamchatka trenches. On the other hand, abundant carbonate is subducted at the Central American margin, where the seafloor is beneath regions of high biological productivity, and at the New Zealand margin, where the seafloor is shallow[38] (Box 2).

Organic carbon is also consumed in the oxic ocean and in the sediments themselves by microbially mediated reactions, so its preservation in sediments requires rapid supply and burial. These conditions are met beneath regions of high biological productivity and in deep-sea fans, where rivers deliver high fluxes of carbon-bearing sediment to the ocean from regions of active uplift and erosion[39]. On the other hand, vast expanses of the ocean are deserts owing to low biological productivity and to the challenge of survival in the harsh oxic sediments[40]. Seafloor currently subducting in the western Pacific spent most of its lifetime traversing the central gyres and is therefore devoid of organic carbon; essentially no organic carbon is subducted at Tonga or Honshu[34]. By contrast, turbidite sediments in the Bengal and Indus fans, which derive from and fringe India's collision zone with Asia, constitute the largest fluxes of organic carbon into trenches. Other margins with notable piles of sediment containing organic carbon include those of Nankai, Cascadia, Alaska, South Chile and the Southern Antilles[34].

Thus, subducting carbon depends on geologic happenstance at locations where a deep-sea fan (high sedimentary organic carbon) or shallow seafloor (high sedimentary carbonate) happen to be near a trench, or where the subducting oceanic plate happens to be created by slow spreading (favouring carbonated serpentinites) or was formed in the Cretaceous period (favouring carbonated oceanic crust). There is exceptionally wide global variation in carbon input to a subduction zone; each downgoing plate has a distinct formation, evolution and sedimentation history (Box 2 and ref. [5]), with potentially large along-strike variations[41]. Global averaging obscures these underlying factors controlling recycling efficiency, which are essential for constructing the long-term history of the deep carbon cycle. This affects not only the amount and distribution of subducting carbon but also its reactivity and isotopic composition, and contributes to heterogeneity in the deep Earth. The former affects the fate of carbon in the subduction zone (discussed in the next section) and the latter serves as a useful tracer for the source of volcanic gases and diamonds (discussed in subsequent sections).

## Reactions in subducting carbon

Once subducting lithosphere and sediment, and the carbon that they carry, are transported beyond the trench (point (I) in Fig. 1), the fate of carbon is determined by a game of subtraction. Rising pressure and temperature transform the subducted materials chemically and

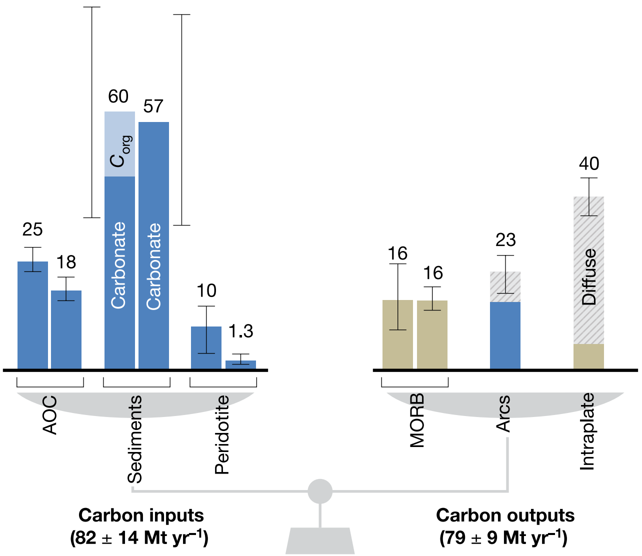
**Box 1**

## Global carbon flux into and out of the mantle

Following the work of Kelemen & Manning[3], several studies have investigated carbon fluxes as part of the Deep Carbon Observatory programme. On the input side, carbon concentration and isotope measurements[30] of AOC have yielded a slightly lower flux (18 Mt C $yr^{-1}$) than that reported in ref. [3] (25 Mt C $yr^{-1}$). Nonetheless, bulk AOC carbon estimates have remained consistent at 400–600 ppm C (refs. [102,105]). Two recent studies[34,106] have revised the C flux in subducting sediments substantially upwards compared to ref. [3] (13–23 Mt C $yr^{-1}$ according to ref. [38]). One estimate (~60 Mt C $yr^{-1}$) includes 20% organic carbon[34] and another (57 Mt C $yr^{-1}$) is based on a model of the calcite compensation depth[106]. Both ref. [3] (10 Mt C $yr^{-1}$) and ref. [23] (1.3 Mt C $yr^{-1}$) estimate low C fluxes in subducting peridotite.

On the output side, identical MORB C fluxes (16 Mt C $yr^{-1}$) are derived from C in vapour-undersaturated volcanic glasses[93] and a coupled degassing model for C and noble gases[107]. The C flux estimates for actively degassing arcs and intraplate volcanoes[10] are immensely improved owing to satellite estimates of S fluxes[7], coupled with recent measurements of the $CO_2$/S ratio in volcanic gases[11]. Considerable diffuse degassing is associated predominantly with intraplate calderas and geothermal systems[10].

On balance, estimates of C input and output fluxes are remarkably similar, although uncertainties are large especially for subducting sedimentary C fluxes. Arc outputs currently represent $27^{+23}_{-16}\%$ of the inputs (based on a Monte Carlo propagation of $2\sigma$ uncertainties), roughly half of that derived from the averages in ref. [3]. This reflects both increases in the sediment input and decreases in the arc output fluxes estimated in recent studies.

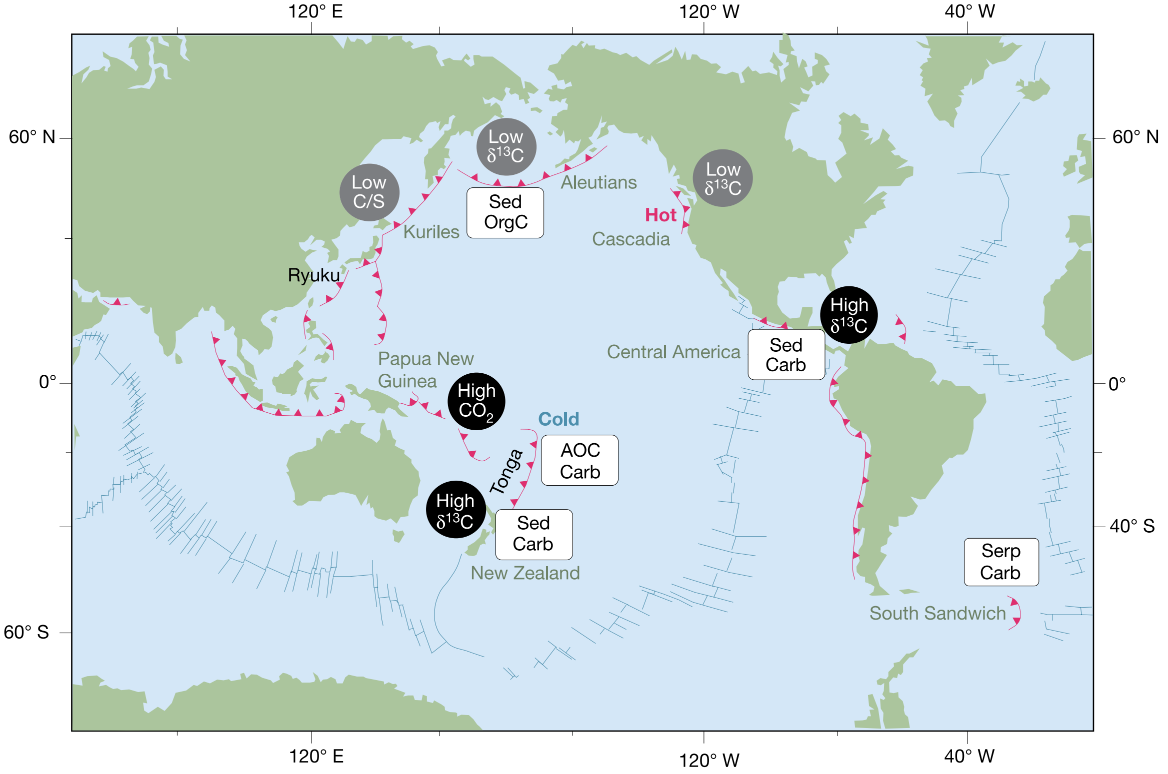


**Box 1 Fig. 1 | Carbon flux balance for inputs to and outputs from the mantle**. All values shown are in megatons per year. Dark blue boxes are specific to subduction zones. Diagonal ruling denotes diffuse degassing. Error bars are $1\sigma$, as quoted or estimated by individual studies. The total input and output fluxes include $1\sigma$ uncertainties calculated by Monte Carlo propagation, but are probably underestimates, given the lack of information on the distribution of values or sources of uncertainties in most studies.

**Box 2**

## Roadmap to carbon subduction and recycling

Each subduction zone is fed by a different feedstock of carbon contained within sediments, oceanic crust and serpentinized mantle of the incoming plate. Sediments may host both organic carbon and carbonate. The subduction zones highlighted in Box 2 Fig. 1 are dominated by one kind of carbon feedstock, and so can serve as sites for natural recycling experiments to determine the downstream fate of different kinds of input. For example, because little sedimentary carbon is subducted at Tonga, AOC carbon recycling can be isolated in the context of cold subduction.



**Box 2 Fig. 1 | Map of subduction systems that isolate different carbon inputs**. Carb, carbonate; OrgC, organic carbon; Sed, sediment; Serp, serpentinite; C/S, $CO_2$/S ratio.

**Box 2 Table | Attributes of each highlighted subduction zone**

| Subduction zone | Attribute | Notes | References |
|---|---|---|---|
| South Sandwich | Carbonated serpentinite | Very slow-spreading crust | 97 |
| Central America | Carbonate sediments | High biological productivity; high $C_{org}$ | 34,38 |
| | High $\delta^{13}C$ in arc | Mean $\delta^{13}C$ of −3.0‰ | 92 |
| Cascadia | Hot slab | Young subducting plate | 98 |
| | Low $\delta^{13}C$ in arc | Mean $\delta^{13}C$ of −8.6‰ | 92 |
| Aleutians | High $C_{org}$ in turbidites | Little carbonate | 34,38 |
| | Low $\delta^{13}C$ in arc | Mean $\delta^{13}C$ of −7.0‰ | 92 |
| Honshu–Kuriles | Low $CO_2$/S in arc | Little sedimentary C | 34,89 |
| Papua New Guinea–Vanuatu | High $CO_2$ flux in arc | Subducting carbonate | 11 |
| Tonga | AOC carbon, cold slab | Little carbon in sediments | 32,38 |
| New Zealand | Carbonate sediments | Shallow seafloor | 34,38 |
| | High $\delta^{13}C$ in arc | Mean $\delta^{13}C$ of −3.0‰ | 92 |

physically, setting the stage for fractional carbon removal by a mix of processes that determine recycling efficiency.

During transit to sub-arc depths, three main processes subtract carbon from the slab: mechanical removal, metamorphic decarbonation and melting (Box 3). Frontal accretion or underplating of sediment during the initial stages of subduction may remove a substantial amount of carbon[41]. The dynamic environment at the top of the slab can result in mixing and removal of slab material. Complex tectonic mixtures of lithologies—mélanges—are important in many exhumed subduction complexes, but to first order they are subject to the same set of processes as their constituent lithologies, and their presence or absence cannot be predicted for modern subduction systems, so they

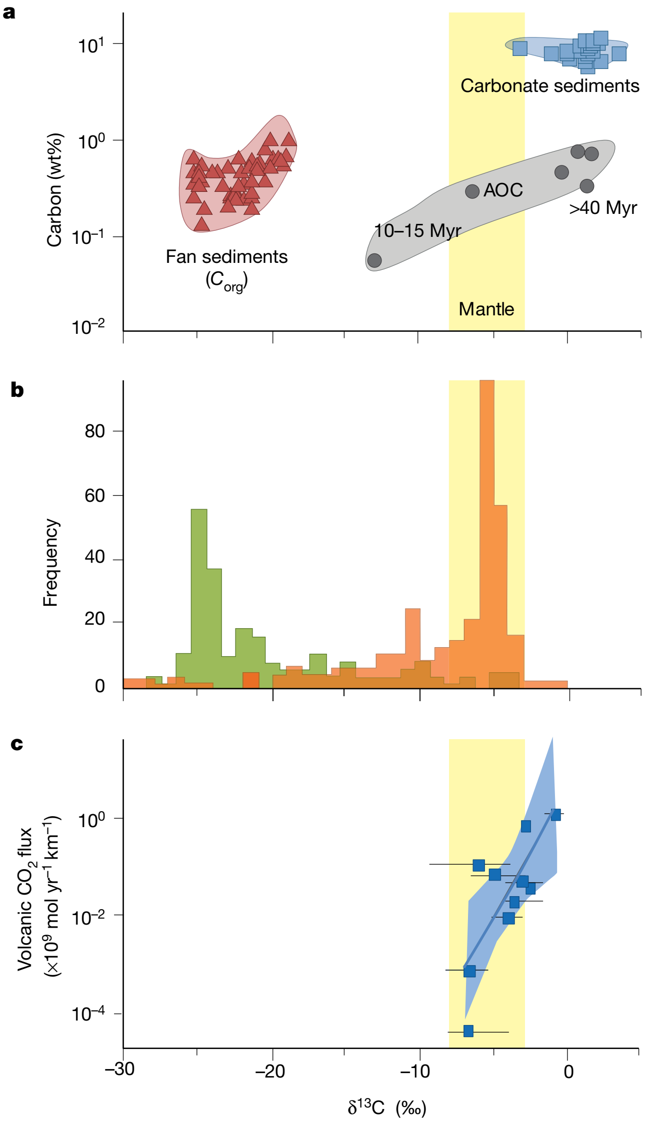


**Fig. 2 | Carbon isotopes in subducted input, diamonds and arc volcanic gas.** a, Subducted carbon input, including that in the upper volcanic layer of AOC older than 10 Myr (refs [30,102]), carbonate sediments from ODP site 765 (ref. [41]), total organic carbon, $C_{org}$, in Bengal Fan sediments[103]. **b**, Diamonds, including 319 eclogitic diamonds (orange; from ref. [30]) and Juina SLD (green; ref. [104]). **c**, Volcanic output per kilometre of arc from ref. [92], where error bars are one standard deviation about the mean value. Yellow shading encompasses mantle values. Diamond populations appear skewed towards light carbon isotopes[74], whereas arc volcanic C is skewed towards high $\delta^{13}C$ values[92]. It is unclear whether this reflects preferential recycling of carbonate to the arc and preferential subduction of more refractory organic carbon to the deeper mantle source of diamonds, or whether ancient diamonds reflect more-reduced sources in Earth's past[14].

are omitted here. Likewise, the extent of subduction erosion of forearc material is debated, as is its carbon content[34,42,43]. Mechanical removal reflects off-scraping and diapiric flow of large rock masses[44,45]. Because this is chiefly evident from past subduction complexes exhumed to the surface in the geologic record, only rarely is it possible to establish where and how mechanical removal operates today. A prominent instructive exception is in the Mariana forearc[46].

The removal of metamorphic carbon from the slab depends on the kindness of water. Carbonate minerals and graphite/diamond are stable on their own up to very high temperatures and pressures; in the absence of any other processes or materials, such carbon is extremely refractory and its recycling efficiency is low. However, subducted lithologies also carry hydrogen and oxygen bound in minerals, and the rising pressure and temperature drive their release to create a free aqueous fluid that is buoyant and reactive with carbon. The fluid strips carbon from these rocks, carries it along its path and deposits it downstream in rocks or melts as a response to changing conditions. Traditional thinking focused only on molecular forms of carbon in the fluid, such as $CO_2$, and concluded that its recycling efficiency is low, except perhaps in the hottest subduction systems[47–49]. Driven by emerging experimental and field studies, recent work also considers the roles of the oxidation state, pH and other rock constituents and dissolved species, generally leading to greater recycling efficiency[3,50–54]. Although carbon solubilities may contribute to substantial redistribution, they are probably insufficient on their own to account for the removal of all carbon in every instance.

The recycling efficiency can be substantially boosted by melting, especially at the slab top. Graphitized organic matter has very low solubility in sediment melts[14], and modern subduction geotherms yield temperatures that are typically too cool to trigger melting of subducted carbonate-bearing peridotite, AOC and sediment at sub-arc depths in the absence of a free water-rich fluid[55–57]. However, water-rich fluids, especially those sourced from dehydrating peridotites at sub-arc depths, can trigger melting, especially in the sedimentary layer. Melts derived from carbonate-bearing AOC and sediment can remove substantial carbon[58,59]. Moreover, in many subduction systems, the slab top beneath the arc is near conditions at which there is complete miscibility between sediment melt and aqueous fluid[60], making continuous dissolution the operative process. Such fluids have great transport capacity, although the role of carbon in them remains poorly understood. Carbonate-rich melts can be produced by the final stages of dehydration of subducted mantle lithosphere beneath the arc front[61].

Any carbon remaining after these processes may be transported in the slab or possibly in the superjacent mantle wedge[3]. At relatively shallow depths the overlying lithospheric mantle is static, but at a certain depth mantle wedge peridotite kinematically couples to the slab and is dragged downwards[62]. Thus, the downward-dragged mantle wedge represents an additional carbon sink created by the subduction process.

Much past debate focused on isolating a single mechanism as the primary pathway for carbon subtraction. Here we emphasize that it is more useful to recognize that several mechanisms can operate in parallel or in series, and that each global segment experiences them in unique proportion and extent. Considering the different carbon inputs to the global subduction system (Box 2), Box 3 describes the operation of the three distinct processing mechanisms for carbon removal. Metamorphic decarbonation is likely to operate in all subduction zones, although the extent to which this removes carbon varies with thermal structure and hydration extent. In the cold Tonga system, it is probably inefficient. The hot Cascadia subduction system stands in stark contrast, with substantial carbon loss by devolatilization/dissolution and melting of predominantly reduced sedimentary carbon. The Central American sediments, rich in biogenic carbon of both oxidized and reduced forms, may fractionate during devolatilization and melting, with preferential recycling of soluble, oxidized carbonate compared to less-soluble organic carbon. The slow-spreading crust feeding the South Sandwich system is expected to be dominated by carbonated peridotite in a young, warm slab. The degree to which mechanical diapiric transport operates today is not known; however, one setting in which it may have been important in the recent past could be the Alpine subduction system. Melting was unlikely as there was no volcanic arc. Though carbon inputs were likely high in Tethyan Mesozoic oceanic crust, the combination of diapirism and dissolution/dehydration could have driven substantial carbon recycling. These examples highlight the conspicuous expressions of individualism in Earth's subduction zones that become obscured by global averaging.

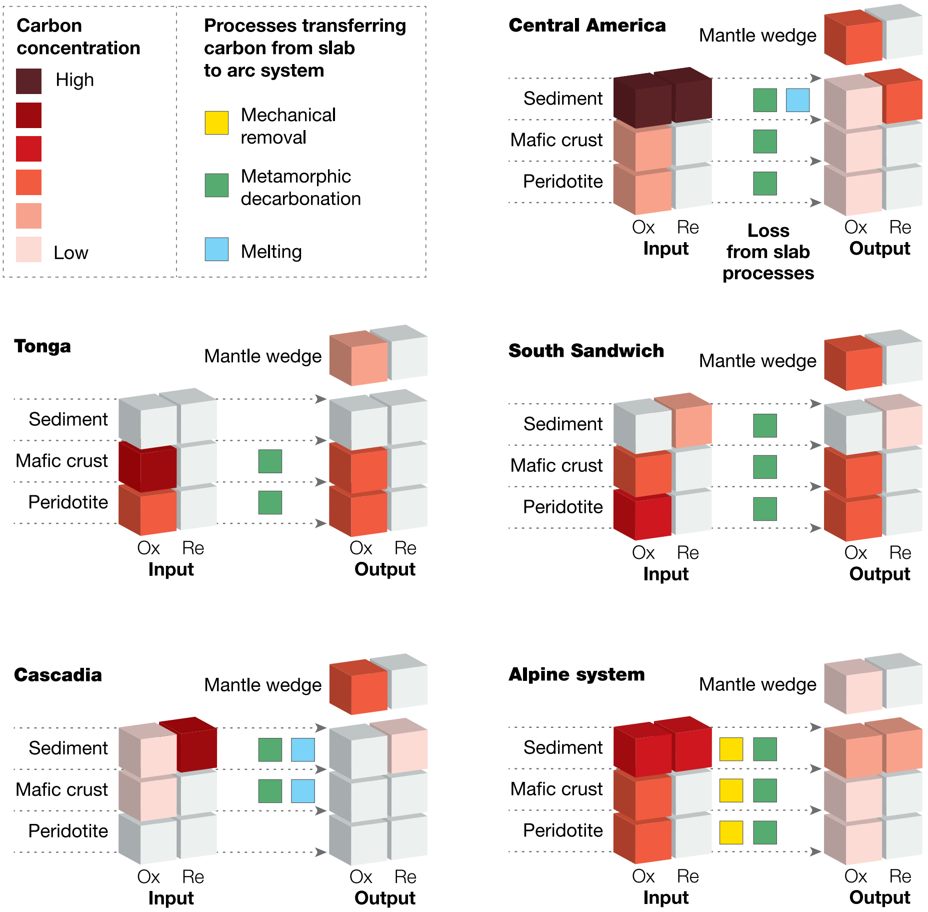
## Carbon beyond the arc: diamonds and more

Carbon that survives beyond sub-arc depths is exceptionally challenging to investigate. In the absence of a fluid phase, even trace carbonate remaining in slab lithologies can trigger melting that may be small in volume but is geochemically important. Most slab geotherms will

**Box 3**

## Carbon removal in the subduction zone

Each subducting system illustrates different mechanisms and recycling efficiencies. Tonga is cool, with little sediment but abundant oxidized carbon in mafic crust of Mesozoic age; the dominant carbon removal process is probably inefficient metamorphic decarbonation. Cascadia has higher temperature, and inputs include considerable amounts of reduced carbon. Combined metamorphic and melting processes probably achieve efficient carbon removal. In Central America there is subduction of substantial oxidized and reduced sedimentary carbon, in a young plate with low carbon concentration in the mafic crust. Geotherms lead to lower slab temperatures than those for Cascadia, but they are sufficient to yield metamorphic and melting loss of sedimentary C, with greater loss of carbonate than less-soluble organic carbon. South Sandwich is a subduction system in which inputs may be dominated by peridotitic carbonate. The Alpine system is shown for a speculative example in which mechanical removal may have been considerable, and carbon input was substantially higher in the Tethyan lithosphere.



**Box 3 Fig. 1 | Schematic representation of carbon inputs, outputs and removal in the global subduction system**. The key (top left) shows vertically arrayed boxes representing lithospheric inputs and outputs in sediment, mafic crust and peridotite, partitioned between oxidized (Ox) and reduced (Re) carbon; outputs include carbon in the mantle wedge. Between the input and output columns are columns illustrating the processes transferring carbon from the slab to the arc system by mechanical removal, metamorphic decarbonation and melting. Examples from subduction segments (from Box 2) show inputs and outputs coloured as in the key in the box. White boxes indicate no carbon.

eventually intersect conditions at which carbonated crust melts[63,64], producing a magma highly enriched in carbonate known as carbonatite, which has extremely high mobility, very similar to that of water[65,66]. Such melts will quickly migrate into the surrounding mantle, where vastly different chemical conditions lead to reactions that transform carbon into new and more resistant forms such as diamond. The durability of diamond allows it to act as a long-term store of information about carbon mixing back into the mantle via subduction.

There are two geodynamically distinct reservoirs of subducted carbon in the mantle: the deep lithospheric keels beneath continents and the more voluminous, convecting upper mantle, transition zone and upper part of the lower mantle. Most diamonds are derived from the volumetrically minor lithosphere; relationships to subducted carbon exist but are difficult to quantify because of the complex history and great age of lithospheric diamonds. However, rare sub-lithospheric diamonds (SLD) appear to be more directly linked to deeper subduction processes extending into the transition zone and beyond. Study of these diamonds is transforming our understanding of the deep carbon cycle.

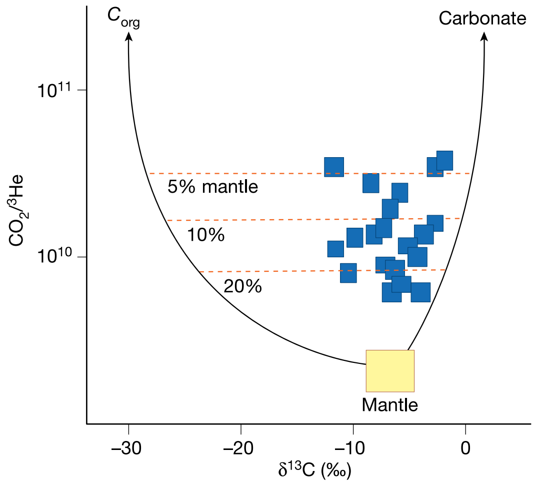
Key to the origin of SLD is the strong chemical contrast between the subducting crust and the ambient convecting mantle. This difference is not just in bulk composition but also, more importantly, in oxidation state. Whereas most crustal rocks are derived from Earth's oxidized

**Box 4**

## Constraints on arc primary magmatic $CO_2$ concentrations and sources

Volcanic-arc gases have $CO_2/^3He$ and $\delta^{13}C$ ratios that are consistent with the mixing of three end-members: MORBs (reflecting background sub-arc mantle), marine carbonate and organic carbon (reflecting subducting sources)[108,109]. Provided that neither the isotopic nor the concentration ratios are fractionated, the proportion of the three end-members in the source carbon can be calculated from the mass balance. A recent compilation of high-temperature gases with minimal crustal contributions[110] found that 80–95% of arc carbon is recycled (Box 4 Fig. 1). This result derives entirely from the higher $CO_2/^3He$ ratios in arc versus MORB gases. Different C and He solubilities and diffusivities during degassing can fractionate this ratio[111], although in MORBs this leads to an average decrease of only 25% in the $CO_2/^3He$ ratio[107]. If the high arc $CO_2/^3He$ ratio reflected preferential He loss, then the arc $CO_2$/S ratio should also be lower than that of MORB (given the lower solubility of $CO_2$ versus S)—it is not[11]. Instead, the combined C–S–He systematics supports the conclusion that most of the $CO_2$ in arc magmas is recycled from the subduction zone, with the $\delta^{13}C$ value reflecting a predominance of carbonate sources.

Box 4 Table combines gas and magma data from two independent approaches and estimates ~1.2 wt% $CO_2$ in primary arc magmas and a $CO_2$ mole fraction of ~0.1 in the aggregate slab fluid/melt (assuming that all C is $CO_2$ and all H is $H_2O$). These quantities, here based on rough averages, have obvious power in testing different slab decarbonation scenarios. The way forward is to compare regional variations in $CO_2/^3He$, $\delta^{13}C$ and $CO_2$/S with different slab inputs (Box 2) and reaction models (Box 3) to quantify the efficiency of the deep carbon cycle.



**Box 4 Fig. 1 | C–He mixing relationships**. Volcanic-arc gases (blue boxes) and mantle-subducting sediment mixing curves from ref. [108]. Dashed lines show the per cent contribution of the mantle end-member.

**Box 4 Table | $CO_2$ in primary (initial, mantle-derived) arc magmas**

| Parameter | Value | Reference/calculation |
|---|---|---|
| 1. Mantle $CO_2$/Nb | 600 | [93] |
| 2. Arc Nb (ppm) | 2 | [99] |
| 3. Arc $CO_2$ from mantle (ppm) | 1,200 | Line 1 × line 2 |
| 4. Mass fraction of mantle end-member in arc $CO_2$ | 0.1 | Box 4 Fig. 1 |
| 5. Primary arc $CO_2$ (wt%) | 1.2 | Line 3/line 4 |
| 6. Arc gas $CO_2$/S | 4 | [11] |
| 7. Arc magma S (ppm) | 3,000 | [100] |
| 8. Primary arc $CO_2$ (wt%) | 1.2 | Line 6 × line 7 |
| 9. Arc magma $H_2O$ (wt%) | 4 | [101] |
| 10. H/C | 1.4 | Line 9/line 5 |
| 11. $CO_2$ mole fraction | 0.11 | From line 10 |

The table lists parameters, values and calculations used to derive estimates for the concentration of $CO_2$ in primary arc magmas. All quantities are expressed in mass units except line 11.

surface reservoirs, the mantle is relatively reducing, and becomes more so with depth. Driven chiefly by changes in the stability of iron-rich minerals with depth[67–71], this leads to conditions that stabilize an iron-rich metallic phase and, at depths greater than about 140 km, produce diamond as the stable form of carbon.

SLD are identified by virtue of the mineral inclusions that they hold, which signal a depth greater than that of the deepest lithospheric keels (>200 km). SLD are younger than lithospheric diamonds, show highly complex growth histories and display strong chemical links to subducted crust[72,73]. Carbon isotope data for SLD are quite variable, extending to very low $\delta^{13}C$ values consistent with carbon derivation from subducted sediment or altered basalt[30,72,74] (Fig. 2b).

If SLD are consistently associated with subducted carbon, then how do they form and how are they transported to the surface? The mineral inclusion assemblages in SLD may show peridotite affinity, but they dominantly host eclogitic mineral assemblages[75]. Traditionally, inclusion 'affinities' are assumed to identify the host rock in which diamond grew. This view implies a narrow set of rock types with a restricted range of chemical characteristics that make it difficult to explain the wide variations observed in SLD. However, an emerging alternative model posits that SLD and their inclusion phases coprecipitate from the fluids and melts derived from subducted crust, so that diamond and its inclusions are products of crystallization in the ambient mantle during reactive flow of melt migrating away from the slab. Low degrees of interaction yield coprecipitating phases that are more similar to the source eclogite, whereas higher degrees of interaction yield an assemblage that is more representative of the ambient mantle. Such a model predicts that inclusion assemblages are chemically linked although they appear to reflect different host rocks, and that there should be straightforward elemental and isotopic patterns consistent with the evolution of conditions from the eclogite source to the ambient mantle. This model better explains the highly enriched character of inclusion minerals, which are otherwise anomalous if they are direct samples of the slab[76]. It also explains C–O stable isotopic systematics[77].

SLD are but a small fraction of diamonds erupted in the kimberlites in which they are found, and diamond-bearing kimberlites are but a small fraction of all kimberlite magmas, which are themselves a minute but important component of global magmatism. As noted earlier, the vast majority of diamonds come from the subcontinental mantle lithosphere. Like SLD, they are best interpreted as metasomatic in origin, and some show evidence for the involvement of subducted carbon, but they differ from SLD in their great age (extending to more than 2.5 Gyr). A wide range of possible mechanisms form lithospheric diamonds, including mixing of mantle and subduction-related fluids, precipitation from metallic melts and mechanical mixing of diamonds formed in subducted slabs with mantle before capture in the lithosphere. As with the happenstance of carbon subduction, the record of subducted carbon found in diamonds is also a highly random sample of material heterogeneously distributed in space and time; it is not in control of its own fate and it depends on largely independent processes to be mined from the deep.

Overall, the total amount of carbon that exists in the form of diamond is unknown. Even if kimberlite eruptions are taken as representative of normal mantle, the overall amount of carbon in diamond is small[3]. It seems likely that the more common return path for carbon stored in metal carbides and diamond is as a component of magmas derived from the mantle; it is typically so chemically modified that little can be said about the processes preceding melting.

## Carbon returned: volcanic gas

Explosive volcanic eruptions—common within the volcanic arcs that form above subduction zones—are driven by the dramatic exsolution of volatiles, including $CO_2$, as magmas rise to the surface. This is due to a strong drop in solubility with decreasing pressure. Magmas can readily dissolve large quantities of carbon as carbonate in the mantle[78], but magmas with more than 1 wt% $CO_2$ will be vapour-saturated and degas throughout their entire ascent in the crust[79]. The standard method used to measure the concentration of magmatic volatiles is to analyse melt inclusions trapped in early-formed minerals. This approach has been successful in estimating the concentrations of $H_2O$, S, Cl and F, but not $CO_2$, owing to its much lower solubility at crustal pressures[80]. Moreover, as much as 90% of the $CO_2$ in a melt inclusion can be exsolved to a shrinkage bubble during cooling[81]. Unfortunately, the literature is replete with measurements of $CO_2$ in melt inclusions compromised by degassing or bubble formation. After bubble reconstruction, arc melt inclusions may contain thousands of parts per million of $CO_2$[82,83], but there is no guarantee that these melts did not lose $CO_2$ by degassing before inclusion formation. Thus, a major outstanding challenge is to determine the $CO_2$ concentration of primary arc magmas (Box 4).

Given that erupted magma has lost its $CO_2$, a more fruitful approach is to measure the lost gas. However, $CO_2$ above the atmospheric background is difficult to measure using remote-sensing techniques. Campaign-style gas measurements show large spatial and temporal variations in $CO_2$ fluxes in different volcanic systems including diffuse degassing from soils, springs and hydrothermal systems[84,85]. Diffuse degassing is generally strongest for large silicic calderas that erupt infrequently[10]. Thus, although satellite detection holds promise for the future[86], estimating $CO_2$ fluxes from volcanoes is difficult owing to its non-steady-state and non-point-source distribution.

A recent promising approach for the estimation of volcanic $CO_2$ fluxes has been to make use of the ratio of $CO_2$ to sulfur. Relatively inexpensive multicomponent gas analyser systems (such as Multi-GAS) have revolutionized our understanding of $CO_2$/S variations at different volcanoes over their eruption cycles. In addition to their great utility in eruption forecasting[87,88], multi-year records from persistently degassing volcanoes show that most gas release occurs passively during the inter-eruptive period[89]. Integrated $CO_2$/S correlates in some volcanic arcs with ratios of non-volatile trace elements (for example, Sr/Nd or Ba/La) linked to subduction recycling of carbonate and slab fluids[9]. Arcs with low $CO_2$/S ratio (for example, in the western Pacific) are associated with a lack of subducting sedimentary carbon, whereas those with higher $CO_2$/S ratios are associated with subducting sedimentary carbonate (Box 2). Sometimes magmas intercept limestones in the upper plate and drive decarbonation[90], a process recognized from an anomalously high $CO_2$/S ratio (>5) that is not linked to other slab tracers[89]. $CO_2$/S data have also led to $CO_2$ flux estimates[11] when coupled to volcanic $SO_2$ fluxes measured from space[7]. The biggest volcanic emitters are found where carbonate-rich sediments subduct (Central America, Vanuatu and Papua New Guinea), whereas low-$CO_2$-flux volcanoes occur where no carbonate sediments subduct (Kuriles-Aleutians; Box 2). These studies show clear signs that volcanic $CO_2$ variations reflect varying subducted input, but longer records are needed to understand temporal variations[91].

Another useful tool is C isotopic ratios in volcanic gases, which may be lower or higher than mantle values for individual volcanoes. Some gases have low $^3He/^4He$ ratios, a signal of upper-crustal contamination[92]; however, the full range of $\delta^{13}C$ values is found in volcanic gases that have mantle-like $^3He/^4He$. Low-$\delta^{13}C$ volcanic carbon occurs at margins where low-$\delta^{13}C$ organic carbon is subducted (for example, the Aleutians), whereas high $\delta^{13}C$ values occur where high-$\delta^{13}C$ carbonate is subducted (for example, Central America; Box 2). Globally, volcanic arcs with the highest C fluxes have the highest $\delta^{13}C$, suggesting a dominance of carbonate sources (Fig. 2c). This could reflect preferential recycling of carbonate in the subduction zone. The $CO_2/^3He$ ratios of arc volcanic gases are consistent with most of the carbon (80–95%) being derived from recycled, subducted sources (Box 4). These constraints, coupled with those from the $CO_2$/S and $CO_2$/Nb ratios, point to primary arc magmas containing up to 1 wt% $CO_2$ (Box 4), an order of magnitude higher than that of average mid-ocean-ridge basalt (MORB; about 1,000 ppm)[93]. Volcanic-arc data are still sparse, however, and portable mass spectrometers mounted on helicopters[94] and field vehicles[95] are providing new methods for carbon isotope analysis in remote regions. Carbon isotope measurements in springs are also revealing subducting and upper-plate sources, as well as sequestration in the biosphere[96].

## Conclusions

Carbon subduction is neither a steady-state nor a globally averaged process. The input fluxes to deep sea trenches are spatially and temporally heterogeneous, as are slab pressure–temperature paths, and thus every subduction zone is a different carbon recycling experiment. In this Review, we have developed three essential viewpoints:

(1) Carbon subduction is geological happenstance. Fixing carbon in oceanic peridotite or burying it in sediments is inherently difficult, as the long path lengths and high solubilities in the deep ocean work against it. Sedimentary carbon is virtually absent from some subduction zones and dominates others (Box 2). Carbonate veining is pervasive in the oceanic crust but its extent varies with crustal age (Fig. 2a). Past (Tethys subduction) or future (Atlantic subduction) carbonate inputs might be larger than those occurring today. At other times, continent collision and erosion deliver organic carbon to deep-sea fans that subduct. Such tectonic events will affect the climate system over long timescales.

(2) Sediments drive the variability in subducting carbon. If the latest flux estimates are accurate (Box 1), then carbonate sediments dominate the subducting flux of carbon by a factor of about two over oceanic-crust and peridotite input fluxes. Recycling of carbonate may explain the high C fluxes and high $\delta^{13}C$ values observed in some volcanic arcs (Fig. 2c). Sediments also supply the greatest flux of organic carbon to the solid Earth (Box 1), and the fate of reduced and isotopically light carbon is critical to the oxidation state of Earth's surface and mantle, and may relate to the low $\delta^{13}C$ values of eclogitic diamonds (Fig. 2b). There may be a complementarity to C isotopes, with heavy carbon isotopes reflecting predominantly carbonate recycling beneath arcs and light carbon isotopes representing deep subduction of organic carbon as a source of diamonds.

(3) There is no single recycling efficiency for carbon. The recycling of carbon back to Earth's surface via arc volcanism depends on the carbon-carrying capacity of fluids and melts in each subduction zone, which in turn are a function of the slab pressure–temperature paths, the pH and redox potentials, and the bulk compositions of each system (Box 3). Not only does the input vary for each subduction zone, but so also do the parameters that govern the recycling efficiency. Current global flux balances (Box 1) place carbon recycling efficiency at 25%, but the associated uncertainties are very large. Even if accurate, this number is insufficiently precise, so the global average carries questionable meaning for any particular margin, and its variation with time

could be substantial. For example, carbonate recycling could be more efficient than organic carbon recycling, and sedimentary carbon could be more easily liberated than oceanic-crust carbon. Modern subduction zones offer many natural experimental settings (Box 2) to test how initial conditions, such as plate age and organic-to-inorganic ratio, affect recycling efficiency (Box 3). Carbon–helium–sulfur systematics point to 80–95% of arc volcanic carbon deriving from the subduction zone (Box 4), so the signal is large. The stakes are high for the climate system, with subducting carbon acting as a source or sink, depending on its recycling efficiency. Coordinated effort is needed on all fronts—quantifying carbon inputs, modelling chemical transfer in the subduction zone and measuring volcanic outputs and deep diamonds—to constrain the global impacts of subducting carbon.

Received: 17 June 2019; Accepted: 11 September 2019;
Published online 16 October 2019.

1. Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time. *Philos. Trans. R. Soc. Lond. B* **361**, 931–950 (2006).
2. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298**, 1–13 (2010).
**This article examines the global carbon budget and the mobility of carbon in the mantle**.
3. Kelemen, P. B. & Manning, C. E. Reevaluating carbon fluxes in subduction zones, what goes down, mostly comes up. *Proc. Natl Acad. Sci. USA* **112**, E3997–E4006 (2015).
**This review summarizes carbon inputs and outputs to the mantle and emphasizes the potential for carbon to be efficiently recycled from the slab and potentially stored in the arc lithosphere**.
4. Hirschmann, M. M. Comparative deep Earth volatile cycles: the case for C recycling from exosphere/mantle fractionation of major ($H_2O$, C, N) volatiles and from $H_2O$/Ce, $CO_2$/Ba, and $CO_2$/Nb exosphere ratios. *Earth Planet. Sci. Lett.* **502**, 262–273 (2018).
**This paper sets carbon recycling in the context of other volatile components**.
5. Galvez, M. E. & Pubellier, M. in *Deep Carbon: Past to Present* (eds Orcutt, B. N. et al.) 276–309 (Cambridge Univ. Press, 2019).
6. Lee, C. T. A., Jiang, H., Dasgupta, R. & Torres, M. in *Deep Carbon: Past to Present* (eds Orcutt, B. N. et al.) 313–357 (Cambridge Univ. Press, 2019).
7. Carn, S. A., Fioletov, V. E., McLinden, C. A., Li, C. & Krotkov, N. A. A decade of global volcanic $SO_2$ emissions measured from space. *Sci. Rep.* **7**, 44095 (2017).
8. Burton, M. R., Sawyer, G. M. & Granieri, D. Deep carbon emissions from volcanoes. *Rev. Mineral. Geochem.* **75**, 323–354 (2013).
9. Aiuppa, A. et al. Gas measurements from the Costa Rica–Nicaragua volcanic segment suggest possible along-arc variations in volcanic gas chemistry. *Earth Planet. Sci. Lett.* **407**, 134–147 (2014).
10. Werner, C. et al. in *Deep Carbon: Past to Present* (eds Orcutt, B. N. et al.) 188–236 (Cambridge Univ. Press, 2019).
11. Aiuppa, A., Fischer, T. P., Plank, T. & Bani, P. $CO_2$ flux emissions from the Earth's most actively degassing volcanoes, 2005–2015. *Sci. Rep.* **9**, 5442 (2019).
12. Friedlingstein, P. et al. Update on $CO_2$ emissions. *Nat. Geosci.* **3**, 811–812 (2010).
13. Hirschmann, M. M. & Dasgupta, R. The H/C ratios of Earth's near-surface and deep reservoirs and consequences for deep Earth volatile cycles. *Chem. Geol.* **262**, 4–16 (2009).
14. Duncan, M. S. & Dasgupta, R. Rise of Earth's atmospheric oxygen controlled by efficient subduction of organic carbon. *Nat. Geosci.* **10**, 387–392 (2017).
15. Evans, K. A. The redox budget of subduction zones. *Earth Sci. Rev.* **113**, 11–32 (2012).
16. Edmond, J. M. & Huh, Y. Non-steady state carbonate recycling and implications for the evolution of atmospheric $PCO_2$. *Earth Planet. Sci. Lett.* **216**, 125–139 (2003).
17. Kent, D. V. & Muttoni, G. Modulation of Late Cretaceous and Cenozoic climate by variable drawdown of atmospheric $pCO_2$ from weathering of basaltic provinces on continents drifting through the equatorial humid belt. *Clim. Past* **9**, 525–546 (2013).
18. Krissansen-Totton, J. & Catling, D. C. Constraining climate sensitivity and continental versus seafloor weathering using an inverse geological carbon cycle model. *Nat. Commun.* **8**, 15423 (2017).
19. Müller, R. D., & Dutkiewicz, A. Oceanic crustal carbon cycle drives 26-million-year atmospheric carbon dioxide periodicities. *Sci Adv.* **4**, eaaq0500 (2018).
20. Volk, T. Sensitivity of climate and atmospheric $CO_2$ to deep-ocean and shallow-ocean carbonate burial. *Nature* **337**, 637–640 (1989).
21. Johnston, F. K., Turchyn, A. V. & Edmonds, M. Decarbonation efficiency in subduction zones: implications for warm Cretaceous climates. *Earth Planet. Sci. Lett.* **303**, 143–152 (2011).
22. Kelemen, P. B. et al. Rates and mechanisms of mineral carbonation in peridotite: natural processes and recipes for enhanced, in situ $CO_2$ capture and storage. *Annu. Rev. Earth Planet. Sci.* **39**, 545–576 (2011).
23. Alt, J. C. et al. The role of serpentinites in cycling of carbon and sulfur: seafloor serpentinization and subduction metamorphism. *Lithos* **178**, 40–54 (2013).
**This review describes the carbon cycle driven by serpentinization and subduction**.
24. Tonarini, S., Leeman, W. P. & Leat, P. T. Subduction erosion of forearc mantle wedge implicated in the genesis of the South Sandwich Island (SSI) arc: evidence from boron isotope systematics. *Earth Planet. Sci. Lett.* **301**, 275–284 (2011).
25. Grevemeyer, I., Ranero, C. R. & Ivandic, M. Structure of oceanic crust and serpentinization at subduction trenches. *Geosphere* **14**, 395–418 (2018).
26. Parai, R. & Mukhopadhyay, S. How large is the subducted water flux? New constraints on mantle regassing rates. *Earth Planet. Sci. Lett.* **317–318**, 396–406 (2012).
27. Cai, C., Wiens, D. A., Shen, W. & Eimer, M. Water input into the Mariana subduction zone estimated from ocean-bottom seismic data. *Nature* **563**, 389 (2018).
28. Naif, S., Key, K., Constable, S. & Evans, R. L. Water-rich bending faults at the Middle America Trench. *Geochem. Geophys. Geosyst.* **16**, 2582–2597 (2015).
29. Shilobreeva, S., Martinez, I., Busigny, V., Agrinier, P. & Laverne, C. Insights into C and H storage in the altered oceanic crust: results from ODP/IODP Hole 1256D. *Geochim. Cosmochim. Acta* **75**, 2237–2255 (2011).
30. Li, K., Li, L., Pearson, D. G. & Stachel, T. Diamond isotope compositions indicate altered igneous oceanic crust dominates deep carbon recycling. *Earth Planet. Sci. Lett.* **516**, 190–201 (2019).
**This paper provides recent data on the carbon content and isotopic composition of AOC and considers the subducted flux and its contribution to diamonds**.
31. Alt, J. C. & Teagle, D. A. H. The uptake of carbon during alteration of ocean crust. *Geochim. Cosmochim. Acta* **63**, 1527–1535 (1999).
32. Gillis, K. M. & Coogan, L. A. Secular variation in carbon uptake into the ocean crust. *Earth Planet. Sci. Lett.* **302**, 385–392 (2011).
33. Hayes, J. M., Strauss, H. & Kaufman, A. J. The abundance of $^{13}C$ in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125 (1999).
34. Clift, P. D. A revised budget for Cenozoic sedimentary carbon subduction. *Rev. Geophys.* **55**, 97–125 (2017).
35. Pälike, H. et al. A Cenozoic record of the equatorial Pacific carbonate compensation depth. *Nature* **488**, 609 (2012).
36. Van Andel, T. H. Mesozoic/Cenozoic calcite compensation depth and the global distribution of calcareous sediments. *Earth Planet. Sci. Lett.* **26**, 187–194 (1975).
37. Syracuse, E. M. & Abers, G. A. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* **7**, Q05017 (2006).
38. Plank, T. in *Treatise on Geochemistry* 2nd edn, Vol. 4 (eds Holland, H. D. & Turekian, K. K.) 607–629 (Elsevier, 2014).
39. Galy, V. et al. Efficient organic carbon burial in the Bengal fan sustained by the Himalayan erosional system. *Nature* **450**, 407–410 (2007).
40. D'Hondt, S. et al. Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. *Nat. Geosci.* **8**, 299–304 (2015).
41. House, B. M., Bebout, G. E. & Hilton, D. R. Carbon cycling at the Sunda margin, Indonesia: a regional study with global implications. *Geology* **47**, 483–486 (2019).
42. Regalla, C., Fisher, D. M., Kirby, E. & Furlong, K. P. Relationship between outer forearc subsidence and plate boundary kinematics along the Northeast Japan convergent margin. *Geochem. Geophys. Geosyst.* **14**, 5227–5243 (2013).
43. Freundt, A. et al. Volatile ($H_2O$, $CO_2$, Cl, S) budget of the Central American subduction zone. *Int. J. Earth Sci.* **103**, 2101–2127 (2014).
44. Gerya, T. V., Stoeckhert, B. & Perchuk, A. L. Exhumation of high-pressure metamorphic rocks in a subduction channel – a numerical simulation. *Tectonics* **21**, 1056 (2002).
45. Gerya, T. V. & Meilick, F. I. Geodynamic regimes of subduction under an active margin: effects of rheological weakening by fluids and melts. *J. Metamorph. Geol.* **29**, 7–31 (2011).
46. Fryer, P., Ambos, E. L. & Hussong, D. M. Origin and emplacement of Mariana forearc seamounts. *Geology* **13**, 774–777 (1985).
47. Kerrick, D. M. & Connolly, J. A. D. Metamorphic devolatilization of subducted marine sediments and the transport of volatiles into the Earth's mantle. *Nature* **411**, 293 (2001).
48. Kerrick, D. M. & Connolly, J. A. D. Metamorphic devolatilization of subducted oceanic metabasalts: implications for seismicity, arc magmatism and volatile recycling. *Earth Planet. Sci. Lett.* **189**, 19–29 (2001).
49. Gorman, P. J., Kerrick, D. M. & Connolly, J. A. D. Modeling open system metamorphic decarbonation of subducting slabs. *Geochem. Geophys. Geosyst.* **7**, Q04007 (2006).
50. Ague, J. J. & Nicolescu, S. Carbon dioxide released from subduction zones by fluid-mediated reactions. *Nat. Geosci.* **7**, 355–360 (2014).
**This paper provides evidence for carbon transport in subduction zone fluids by dissolution of $CaCO_3$ versus metamorphic decarbonation**.
51. Galvez, M. E., Connolly, J. A. D. & Manning, C. E. Implications for metal and volatile cycles from the pH of subduction zone fluids. *Nature* **539**, 420–424 (2016).
52. Galvez, M. E., Manning, C. E., Connolly, J. A. D. & Rumble, D. The solubility of rocks in metamorphic fluids: a model for rock-dominated conditions to upper mantle pressure and temperature. *Earth Planet. Sci. Lett.* **430**, 486–498 (2015).
53. Gorce, J. S., Caddick, M. J. & Bodnar, R. J. Thermodynamic constraints on carbonate stability and carbon volatility during subduction. *Earth Planet. Sci. Lett.* **519**, 213–222 (2019).
54. Sverjensky, D. A. & Huang, F. Diamond formation due to a pH drop during fluid–rock interactions. *Nat. Commun.* **6**, 8702 (2015).

55. Tsuno, K. & Dasgupta, R. Melting phase relation of nominally anhydrous, carbonated pelitic-eclogite at 2.5–3.0 GPa and deep cycling of sedimentary carbon. *Contrib. Mineral. Petrol*. **161**, 743–763 (2011).
56. Dasgupta, R., Hirschmann, M. M. & Withers, A. C. Deep global cycling of carbon constrained by the solidus of anhydrous, carbonated eclogite under upper mantle conditions. *Earth Planet. Sci. Lett*. **227**, 73–85 (2004).
57. Hammouda, T. & Keshav, S. Melting in the mantle in the presence of carbon: review of experiments and discussion on the origin of carbonatites. *Chem. Geol*. **418**, 171–188 (2015).
58. Grassi, D. & Schmidt, M. W. The melting of carbonated pelites from 70 to 700 km depth. *J. Petrol*. **52**, 765–789 (2011).
59. Martin, L. A. & Hermann, J. Experimental phase relations in altered oceanic crust: implications for carbon recycling at subduction zones. *J. Petrol*. **59**, 299–320 (2018).
60. Cooper, L. B. et al. Global variations in $H_2O$/Ce: 1. Slab surface temperatures beneath volcanic arcs. *Geochem. Geophys. Geosyst*. **13**, Q03024 (2012).
61. Poli, S. Carbon mobilized at shallow depths in subduction zones by carbonatitic liquids. *Nat. Geosci*. **8**, 633–636 (2015).
62. van Keken, P. E., Kiefer, B. & Peacock, S. M. High resolution models of subduction zones: implications for mineral dehydration reactions and the transport of water into the deep mantle. *Geochem. Geophys. Geosyst*. **3**, 1056 (2002).
63. Thomson, A. R., Walter, M. J., Kohn, S. C. & Brooker, R. A. Slab melting as a barrier to deep carbon subduction. *Nature* **529**, 76–79 (2016).
**This paper presents new results on slab melting below the mantle transition zone, which support a new reaction–transport model of sub-lithospheric diamond formation**.
64. Sun, C. & Dasgupta, R. Slab–mantle interaction, carbon transport, and kimberlite generation in the deep upper mantle. *Earth Planet. Sci. Lett*. **506**, 38–52 (2019).
65. Kono, Y. et al. Ultralow viscosity of carbonate melts at high pressures. *Nat. Commun*. **5**, 5091 (2014).
66. Stagno, V. et al. in *Carbon in Earth* (eds Manning, C. E. et al.) (American Geophysical Union, 2019).
67. O'Neill, H. S. C., Rubie, D. C., Canil, D., Geiger, C. & Ross, C. R. in *Evolution of the Earth and Planets* Vol. 74 (eds Takahashi, E. et al.) 73–88 (American Geophysical Union, Washington, 1993).
68. Rohrbach, A. et al. Metal saturation in the upper mantle. *Nature* **449**, 456–458 (2007).
69. Frost, D. J. et al. Experimental evidence for the existence of iron-rich metal in Earth's lower mantle. *Nature* **428**, 409–412 (2004).
70. Frost, D. J. & McCammon, C. A. The redox state of Earth's mantle. *Annu. Rev. Earth Planet. Sci*. **36**, 389–420 (2008).
71. Ballhaus, C. Is the upper mantle metal-saturated? Earth Planet. *Sci. Lett*. **132**, 75–86 (1995).
72. Shirey, S. B. et al. Diamonds and the geology of mantle carbon. *Rev. Mineral. Geochem*. **75**, 355–421 (2013).
73. Smith, E. M. et al. Blue boron-bearing diamonds from Earth's lower mantle. *Nature* **560**, 84 (2018).
74. Cartigny, P. Stable isotopes and the origin of diamond. *Elements* **1**, 79–84 (2005).
75. Shirey, S. B. et al. Diamonds and the mantle geodynamics of carbon: deep mantle carbon evolution from the diamond record. In *Deep Carbon: Past to Present* (eds Orcutt, B. N. et al.) 89–128 (Cambridge Univ. Press, 2019).
**This review provides the most up-to-date information on diamond formation and links to subduction**.
76. Thomson, A. R. et al. Trace element composition of silicate inclusions in sub-lithospheric diamonds from the Juina-5 kimberlite: evidence for diamond growth from slab melts. *Lithos* **265**, 108–124 (2016b).
77. Burnham, A. D. et al. Stable isotope evidence for crustal recycling as recorded by superdeep diamonds. *Earth Planet. Sci. Lett*. **432**, 374–380 (2015).
78. Dasgupta, R. et al. Carbon-dioxide-rich silicate melt in the Earth's upper mantle. *Nature* **493**, 211–215 (2013).
79. Wallace, P. J. Volatiles in subduction zone magmas: concentrations and fluxes based on melt inclusion and volcanic gas data. *J. Volcanol. Geotherm. Res*. **140**, 217–240 (2005).
80. Wallace, P., Plank, T., Edmonds, M. & Hauri, E. H. in *The Encyclopedia of Volcanoes* 2nd edn Vol. I (eds. Sigurdsson H.) 163–183 (Elsevier, 2015).
81. Moore, L. R. et al. Bubbles matter: an assessment of the contribution of vapor bubbles to melt inclusion volatile budgets. *Am. Mineral*. **100**, 806–823 (2015).
82. Mironov, N. et al. Quantification of the $CO_2$ budget and $H_2O$–$CO_2$ systematics in subduction-zone magmas through the experimental hydration of melt inclusions in olivine at high $H_2O$ pressure. *Earth Planet. Sci. Lett*. **425**, 1–11 (2015).
83. Aster, E. M. et al. Reconstructing $CO_2$ concentrations in basaltic melt inclusions using Raman analysis of vapor bubbles. *J. Volcanol. Geotherm. Res*. **323**, 148–162 (2016).
84. James, E. R., Manga, M. & Rose, T. P. $CO_2$ degassing in the Oregon Cascades. *Geology* **27**, 823–826 (1999).
85. Chiodini, G. et al. Carbon dioxide diffuse degassing and estimation of heat release from volcanic and hydrothermal systems. *J. Geophys. Res*. **110**, B08204 (2005).
86. Schwandner, F. M. et al. Spaceborne detection of localized carbon dioxide sources. *Science* **358**, eaam5782 (2017).
87. Aiuppa, A. et al. Patterns in the recent 2007–2008 activity of Mount Etna volcano investigated by integrated geophysical and geochemical observations. *Geochem. Geophys. Geosyst*. **11**, Q09008 (2010).
88. de Moor, J. M. et al. Turmoil at Turrialba Volcano (Costa Rica): degassing and eruptive processes inferred from high-frequency gas monitoring. *J. Geophys. Res*. **121**, 5761–5775 (2016).
89. Aiuppa, A., Fischer, T. P., Plank, T., Robidoux, P. & Di Napoli, R. Along-arc, inter-arc and arc-to-arc variations in volcanic gas $CO_2/S_T$ ratios reveal dual source of carbon in arc volcanism. *Earth Sci. Rev*. **168**, 24–47 (2017).
**This paper provides an example of the potential power of MultiGas data from volcanic vents to constrain $CO_2$ sources in the subduction zone**.
90. Lee, C.-T. A. et al. Continental arc–island arc fluctuations, growth of crustal carbonates, and long-term climate change. *Geosphere* **9**, 21–36 (2013).
91. de Moor, J. M. et al. A new sulfur and carbon degassing inventory for the Southern Central American Volcanic Arc: the importance of accurate time-series data sets and possible tectonic processes responsible for temporal variations in arc-scale volatile emissions. *Geochem. Geophys. Geosyst*. **18**, 4437–4468 (2017).
92. Mason, E., Edmonds, M. & Turchyn, A. V. Remobilization of crustal carbon may dominate volcanic arc emissions. *Science* **357**, 290–294 (2017).
**This paper reviews carbon isotopic compositions for volcanic gases globally and discusses crustal versus subducted sources of volcanic carbon and the implications for the carbon cycle**.
93. Le Voyer, M. et al. Carbon fluxes and primary magma $CO_2$ contents along the global mid-ocean ridge system. *Geochem. Geophys. Geosyst*. **20**, 1387–1424 (2019).
94. Fischer, T. P. & Lopez, T. M. First airborne samples of a volcanic plume for $\delta^{13}C$ of $CO_2$ determinations. *Geophys. Res. Lett*. **43**, 3272–3279 (2016).
95. Rizzo, A. L. et al. Real-time measurements of the concentration and isotope composition of atmospheric and volcanic $CO_2$ at Mount Etna (Italy). *Geophys. Res. Lett*. **41**, 2382–2389 (2014).
96. Barry, P. H. et al. Forearc carbon sink reduces long-term volatile recycling into the mantle. *Nature* **568**, 487–492 (2019); correction **568**, E7 (2019).
97. Lawver, L. A. & Dick, H. J. The American-Antarctic Ridge. *J. Geophys. Res*. **88**, 8193–8202 (1983).
98. Syracuse, E. M., van Keken, P. E. & Abers, G. A. The global range of subduction zone thermal models. *Phys. Earth Planet. Inter*. **183**, 73–90 (2010).
99. Turner, S. J. & Langmuir, C. H. What processes control the chemical compositions of arc front stratovolcanoes? Geochem. *Geophys. Geosyst*. **16**, 1865–1893 (2015).
100. Zimmer, M. M. et al. The role of water in generating the calc-alkaline trend: new volatile data for Aleutian magmas and a new tholeiitic index. *J. Petrol*. **51**, 2411–2444 (2010).
101. Plank, T., Kelley, K. A., Zimmer, M. M., Hauri, E. H. & Wallace, P. J. Why do mafic arc magmas contain $\sim$4 wt% water on average? *Earth Planet. Sci. Lett*. **364**, 168–179 (2013).
102. Alt, J. C. Stable isotopic composition of upper oceanic crust formed at a fast spreading ridge, ODP Site 801. *Geochem. Geophys. Geosyst*. **4**, 8908 (2003).
103. Galy, V., France-Lanord, C., Peucker-Ehrenbrink, B. & Huyghe, P. Sr–Nd–Os evidence for a stable erosion regime in the Himalaya during the past 12 Myr. *Earth Planet. Sci. Lett*. **290**, 474–480 (2010).
104. Thomson, A. R. et al. Origin of sub-lithospheric diamonds from the Juina-5 kimberlite (Brazil): constraints from carbon isotopes and inclusion compositions. *Contrib. Mineral. Petrol*. **168**, 1081 (2014).
105. Jarrard, R. D. Subduction fluxes of water, carbon dioxide, chlorine, and potassium. *Geochem. Geophys. Geosyst*. **4**, 8905 (2003).
106. Dutkiewicz, A., Müller, R. D., Cannon, J., Vaughan, S. & Zahirovic, S. Sequestration and subduction of deep-sea carbonate in the global ocean since the Early Cretaceous. *Geology* **47**, 91–94 (2018).
107. Tucker, J. M., Mukhopadhyay, S. & Gonnermann, H. M. Reconstructing mantle carbon and noble gas contents from degassed mid-ocean ridge basalts. *Earth Planet. Sci. Lett*. **496**, 108–119 (2018).
108. Sano, Y. & Marty, B. Origin of carbon in fumarolic gas from island arcs. *Chem. Geol*. **119**, 265–274 (1995).
109. Hilton, D. R., Fischer, T. P. & Marty, B. Noble gases and volatile recycling at subduction zones. *Rev. Mineral. Geochem*. **47**, 319–370 (2002).
110. Kagoshima, T. et al. Sulphur geodynamic cycle. *Sci. Rep*. **5**, 8330 (2015).
111. Gonnermann, H. M. & Mukhopadhyay, S. Non-equilibrium degassing and a primordial source for helium in ocean-island volcanism. *Nature* **449**, 1037–1040 (2007).

**Acknowledgements** This work was supported by the Deep Carbon Observatory. We thank the late E. Hauri, whose leadership and vision put into motion much of the research on carbon reservoirs and fluxes and on volcanic gas and magma chemistry reported here. We thank A. Malinverno, A. Thomson, S. Shirey, O. Tschauener, M. Walter, A. Aiuppa, T. Fischer, E. Cottrell and D. Kent for discussions, and J. M. de Moor and D. Muller for their comments.

**Author contributions** Both T.P. and C.E.M. contributed to the structure, content and message of this review.

**Competing interests** The authors declare no competing interests.

**Additional information**

**Reprints and permissions information** is available at http://www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to T.P.

**Peer review information** *Nature* thanks J. Maarten de Moor, Dietmar Muller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ARTICLE

OPEN

https://doi.org/10.1038/s41586-019-1545-0

# Mapping 123 million neonatal, infant and child deaths between 2000 and 2017

A list of authors and their affiliations appears in the online version of the paper.

**Since 2000, many countries have achieved considerable success in improving child survival, but localized progress remains unclear. To inform efforts towards United Nations Sustainable Development Goal 3.2—to end preventable child deaths by 2030—we need consistently estimated data at the subnational level regarding child mortality rates and trends. Here we quantified, for the period 2000–2017, the subnational variation in mortality rates and number of deaths of neonates, infants and children under 5 years of age within 99 low- and middle-income countries using a geostatistical survival model. We estimated that 32% of children under 5 in these countries lived in districts that had attained rates of 25 or fewer child deaths per 1,000 live births by 2017, and that 58% of child deaths between 2000 and 2017 in these countries could have been averted in the absence of geographical inequality. This study enables the identification of high-mortality clusters, patterns of progress and geographical inequalities to inform appropriate investments and implementations that will help to improve the health of all populations.**

Gains in child survival have long served as an important proxy measure for improvements in overall population health and development[1,2]. Global progress in reducing child deaths has been heralded as one of the greatest success stories of global health[3]. The annual global number of deaths of children under 5 years of age (under 5)[4] has declined from 19.6 million in 1950 to 5.4 million in 2017. Nevertheless, these advances in child survival have been far from universally achieved, particularly in low- and middle-income countries (LMICs)[4]. Previous subnational child mortality assessments at the first (that is, states or provinces) or second (that is, districts or counties) administrative level indicate that extensive geographical inequalities persist[5–7].

Progress in child survival also diverges across age groups[4]. Global reductions in mortality rates of children under 5—that is, the under-5 mortality rate (U5MR)—among post-neonatal age groups are greater than those for mortality of neonates (0–28 days)[4,8]. It is relatively unclear how these age patterns are shifting at a more local scale, posing challenges to ensuring child survival. To pursue the ambitious Sustainable Development Goal (SDG) of the United Nations[9] to "end preventable deaths of newborns and children under 5" by 2030, it is vital for decision-makers at all levels to better understand where, and at what ages, child survival remains most tenuous.

## Precision public health and child mortality

Country-level estimates facilitate international comparisons but mask important geographical heterogeneity. Previous assessments of mortality of children under 5 have noted significant within-country heterogeneity, particularly in sub-Saharan Africa[5,7,10–14], as well as in Brazil[15], Iran[16] and China[17]. Understanding public health risks at more granular subpopulation levels is central to the emerging concept of precision public health[18], which uses "the best available data to target more effectively and efficiently interventions…to those most in need"[18]. Efforts to produce high-resolution estimates of mortality of children under 5, determinants at scales that cover the multiple countries are emerging, including for vaccine coverage[19,20], malaria[21], diarrhoea[22] and child growth failure[23,24]. In a previous study, we produced comprehensive estimates of African child mortality rates at a 5 × 5-km scale for 5-year intervals[5]. For areas outside of Africa, in which 72% of the world's children live and 46% of global child deaths occurred in 2017[4], subnational heterogeneity remains mostly undescribed[25].

Here we produce estimates of death counts and mortality rates of children under 5, infants (under 1 years of age) and neonates (0–28 days) in 99 countries at policy-relevant subnational scales (first and second administrative levels) for each year from 2000 to 2017. We fit a geostatistical discrete hazards model to a large dataset that is composed of 467 geo-referenced household surveys and censuses, representing approximately 15.9 million births and 1.1 million deaths of children from 2000 to 2017. Our model includes socioeconomic, environmental and health-related spatial covariates with known associations to child mortality and uses a Gaussian process random effect to exploit the correlation between data points near each other across dimensions of space, time and age group, which helps to mitigate the limitations associated with data sparsity in our estimations. For this study, we report U5MR as the expected number of deaths per 1,000 live births, reflecting the probability of dying before the age of 5 for a given location and year.

## Unequal rates of child mortality

The risk of a newborn dying before their fifth birthday varies tremendously based on where in the world, and within their country, they are born. Across the 99 countries in this study, we estimate that U5MR varied as much as 24-fold at the national level in 2017, with the highest rate in the Central African Republic of 123.9 deaths (95% uncertainty interval, 104.9–148.2) per 1,000 live births, and the lowest rate in Cuba of 5.1 deaths (4.4–6.0)[4]. We observed large subnational variation within countries in which overall U5MR was either high or comparatively low. For example, in Vietnam, rates across second administrative units (henceforth referred to as 'units') varied 5.7-fold, from 6.9 (4.6–9.8) in the Tenth District in Hồ Chí Minh City to 39.7 (28.1–55.6) in Mường Tè District in the Northwest region (Figs. 1b, 2).

Decreases in U5MR between 2000 and 2017 were evident to some extent throughout all units (Figs. 1a, b, 2). No unit showed a significant increase in U5MR in this period, and in most units U5MR decreased greatly, even in units in which the mortality risk was the highest. Out of 17,554 units, 60.3% (10,585 units) showed a significant (defined as 95% uncertainty intervals that did not overlap) decrease in U5MR between 2000 and 2017. Across units in 2000, U5MR ranged from 7.5 (5.0–10.6) in Santa Clara district, Villa Clara province, Cuba, to 308.4 (274.9–348.4) in the Sabon Birni Local Government Area of Sokoto State, Nigeria. By 2017, the unit with the highest estimated U5MR across all 99 countries was Garki Local Government Area, Jigawa state, Nigeria, at 195.1 (158.6–230.9). Overall, the total percentage of units with a U5MR higher than 80 deaths per 1,000 live births decreased from 28.9% (5,070) of units in 2000 to 7.0% (1,236) in 2017. Furthermore,

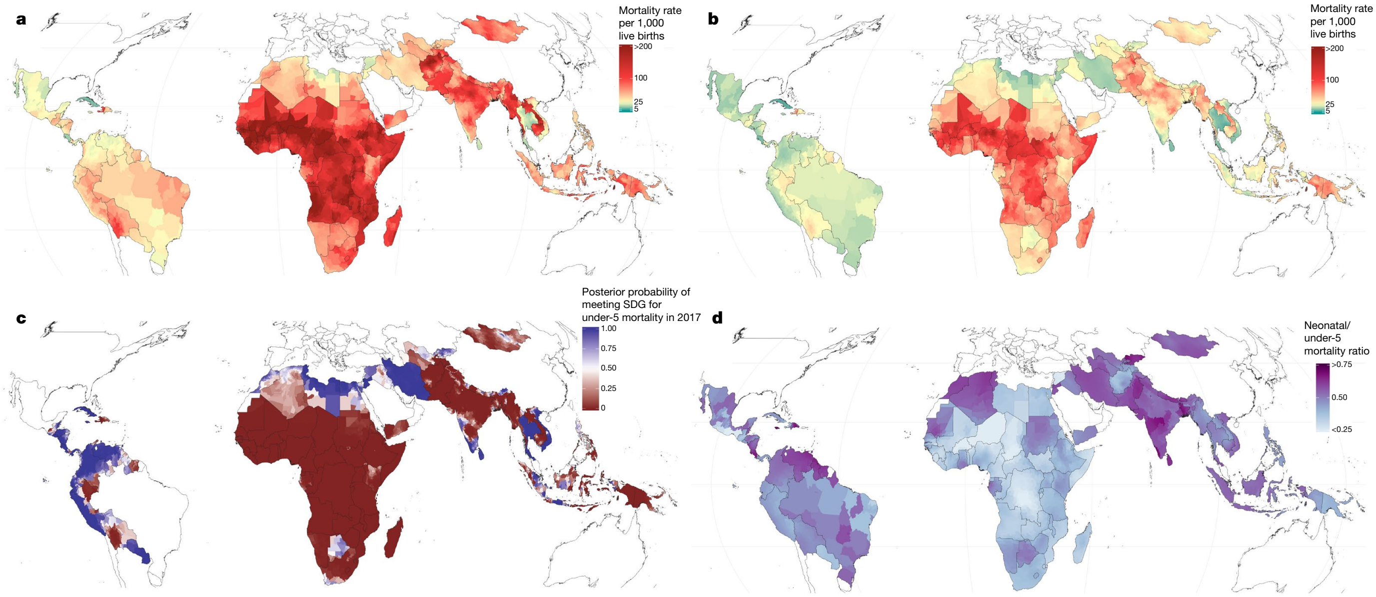


**Fig. 1 | U5MR estimates in 99 LMICs. a**, U5MR at the second administrative level in 2000. **b**, U5MR at the second administrative level in 2017. **c**, Modelled posterior exceedance probability that a given second administrative unit had achieved the SDG 3.2 target of 25 deaths per 1,000 live births for children under 5 in 2017. **d**, Proportion of mortality of children under 5 occurring in the neonatal (0–28 days) group at the second administrative level in 2017.

32% of units, representing 11.9% of the under-5 population in the 99 countries, had already met SDG 3.2 for U5MR with a 90% certainty threshold (Fig. 1c). For neonatal mortality, 34% of units met the target of $\leq$12 deaths per 1,000 live births (Extended Data Fig. 1). Within countries, successes were mixed in some cases. For example, Colombia, Guatemala, Libya, Panama, Peru and Vietnam had all achieved SDG 3.2 for U5MR at the national level by 2017, but each country had units that did not achieve the goal with 90% certainty (Fig. 1c).

Successful reductions in child mortality were also observed throughout entire countries. For example, in 43 LMICs across several world regions, the worst-performing unit in 2017 had a U5MR that was lower than the best-performing unit in 2000 (Fig. 2). Nearly half of these countries were in sub-Saharan Africa. Rwanda showed notable progress during the study period, reducing mortality from 144.0 (130.0–161.6) in its best-achieving district in 2000 (Rubavu) to 57.2 (47.4–72.1) in its worst-achieving district in 2017 (Kayonza). These broad reductions in U5MR have also led to a convergence of absolute subnational geographical inequalities, although relative subnational inequalities appear to be mostly unchanged between 2000 and 2017 (Fig. 2 and Supplementary Fig. 6.12). Despite this success, the highest U5MRs in 2017 were still largely concentrated in areas in which rates were highest in 2000 (Fig. 1a, b). We observed estimated U5MR $\geq$ 80 across large geographical areas in Western and Central sub-Saharan Africa, and within Afghanistan, Cambodia, Haiti, Laos and Myanmar (Fig. 1b).

Deaths of neonates (0–28 days of age) and post-neonates (28–364 days of age) have come to encompass a larger fraction of overall mortality of children under 5 in recent years. By 2017 (Fig. 1d), neonatal mortality increased as a proportion of total deaths of children under 5 in 91% (90) of countries and for 83% (14,656) of units compared to 2000. In almost all places where U5MR decreased, the share of the mortality burden increased in the groups of children with younger ages. Similarly, the mortality of infants (<1 year) has increased relative to the mortality for children who are 1–4 years of age in many areas. For example, in the Diourbel Region, Senegal, infant mortality constituted 54.4% (52.4–56.6) of total mortality of children under 5 in 2000; by 2017, the relative contribution of infant mortality was 73.2% (70.3–75.8). This shift towards mortality predominantly affecting neonates and infants was not as evident in all locations; mortality for children aged 1–4 years was responsible for more than 30% of overall under-5 deaths in 13% (2,226) of units, mostly within high-mortality areas in sub-Saharan Africa.

## Distribution of under-5 deaths may not follow rates

The goal of mortality-reduction efforts is ultimately to prevent premature deaths, and not just to reduce mortality rates. Across the countries studied here, there were 3.5 million (41%) fewer deaths of children under 5 in 2017 than in 2000 (5.0 million compared to 8.5 million). At the national level, the largest number of child deaths in 2017 occurred in India (1.04 (0.98–1.10) million), Nigeria (0.79 (0.65–0.96) million), Pakistan (0.34 (0.27–0.41) million) and the Democratic Republic of the Congo (0.25 (0.21–0.31) million) (Fig. 3a). Within these countries, the geographical concentration of the deaths of the children varied. In Pakistan, over 50% of child deaths in 2017 occurred in Punjab province, which had a U5MR of 63.3 (54.1–76.0) deaths per 1,000 live births (Fig. 3b). By contrast, 50% of child deaths in the Democratic Republic of the Congo in 2017 occurred across 9 out of 26 provinces. Such findings are in a large part artefacts of how borders are drawn around various at-risk populations (the provinces above account for 53% and 63%, respectively, of the under-5 population that is at risk in these two countries), but can have a real impact at the level at which planning occurs. Some concentrated areas with apparent high absolute numbers of deaths highlighted by local-level estimates become less noticeable when reporting at aggregated administrative levels; for example, areas across Guatemala, Honduras and El Salvador are visually striking hotspots in Fig. 3d, but less so in Fig. 3b, c.

Our estimates indicate that targeting areas with a ‘high’ U5MR of 80 will have a lower overall effect than in previous years owing to the reductions in mortality rates. In 2000, 23.7% of child deaths—representing 2.0 (1.7–2.4) million deaths—occurred in regions in which U5MR was less than 80 that year (Fig. 4). By comparison, in 2017, 69.5% of child deaths occurred in areas in which U5MR was below 80. A growing proportion of deaths of children under 5 are occurring in ‘low’-mortality areas; 7.3% (5.1–10.2) of all deaths of children under 5 in 2017 occurred in locations in which the U5MR was below the SDG 3.2 target rate of 25, compared to 1.2% (0.9–1.6) in 2000. For instance, Lima, Peru, has a U5MR in the 8th percentile of units in this study, yet it ranks in the 96th percentile of highest number of deaths of children under 5.

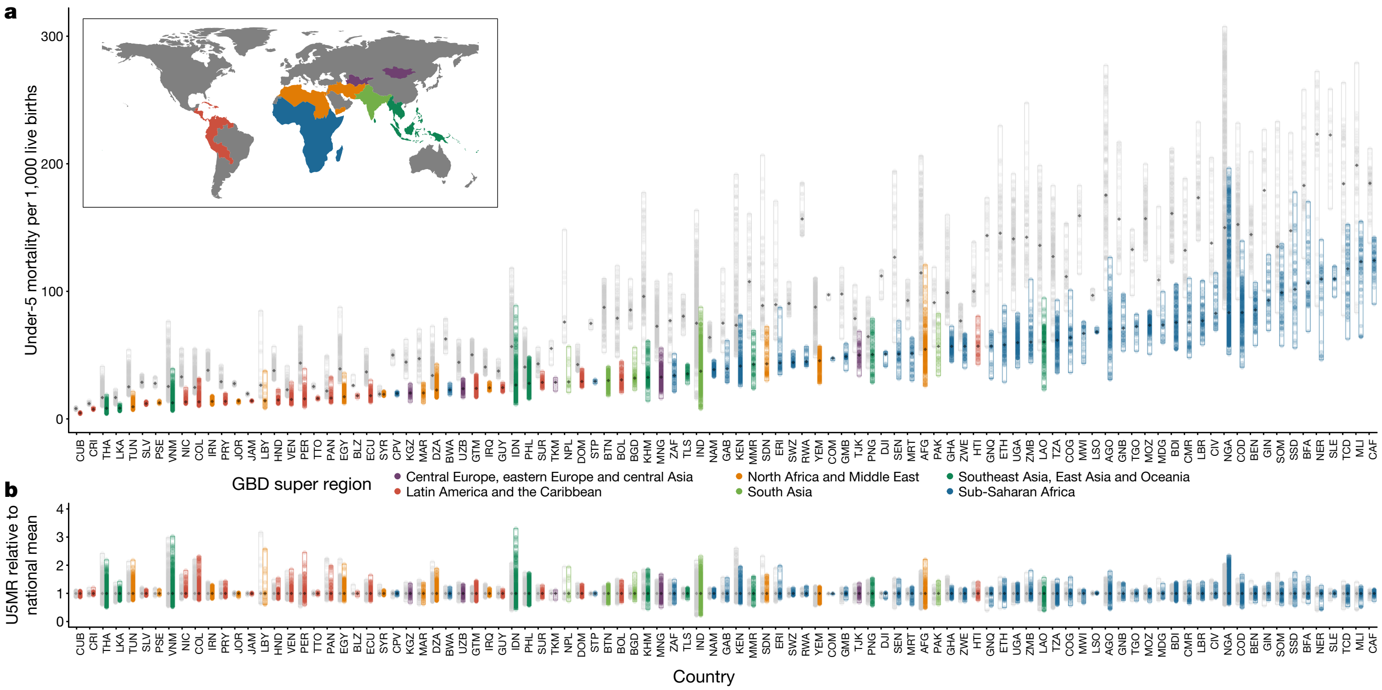


**Fig. 2 | Geographical inequality in U5MR across 99 countries for 2000 and 2017. a**, Absolute inequalities. Range of U5MR estimates in second administrative-level units across 99 LMICs. **b**, Relative inequalities. Range of ratios of U5MR estimates in second administrative-level units relative to country means. Each dot represents a second administrative-level unit. The lower bound of each bar represents the second administrative-level unit with the lowest U5MR in each country. The upper end of each bar represents the second administrative-level unit with the highest U5MR in each country. Thus, each bar represents the extent of geographical inequality in U5MRs estimated for each country. Bars indicating the range in 2017 are coloured according to their Global Burden of Disease super-region. Grey bars indicate the range in U5MR in 2000. The diamond in each bar represents the median U5MR estimated across second administrative-level units in each country and year. A coloured bar that is shorter than its grey counterpart indicates that geographical inequality has narrowed.

Despite population growth, child deaths have declined due to the outpaced decline in U5MR. For example, there were a total of 8.5 (7.2–10.0) million deaths of children under 5 in the countries in this study in 2000; had the 2017 under-5 population been exposed to the same U5MRs that were observed in 2000, there would have been 10.6 (9.0–12.5) million deaths in 2017. Instead, we observed 5.0 (3.8–6.6) million deaths in 2017 (Extended Data Fig. 5).

Finally, we combine estimates of subnational variation in mortality rates and populations to gain a better understanding of the impact of geographical inequality. Overall, 2.7 (2.5–2.9) million deaths, or 54% of the total number of deaths of children under 5, would have been averted in 2017 had all units had a U5MR that matched the best-performing unit in each respective country (Extended Data Fig. 2). Over the 2000–2017 period, this number is 71.8 (68.5–74.9) million deaths, or 58% (55–61) of the total number of deaths of children under 5. Total deaths attributable to inequality in this scenario ranged from 13 (6–24) deaths in Belize to 0.84 (0.72–0.99) million deaths in India. Furthermore, had all units met the SDG 3.2 target of 25 deaths per 1,000, an estimated 2.6 (2.3–2.8) million deaths of children under 5 would have been averted in 2017.

## Discussion

This study offers a comprehensive, geospatially resolved resource for national and subnational estimates of child deaths and mortality rates for 99 LMICs, where 93% of the world's child deaths[4] occurred in 2017. Gains in child survival varied substantially within the vast majority of countries from 2000 to 2017. Countries such as Vietnam, for example, showed more than fivefold variation in mortality rates across second administrative-level units. The inconsistency of successes, even at subnational levels, indicates how differences in health policy, financial resources, access to and use of health services, infrastructure, and economic development ultimately contribute to millions of lives cut short[25–27]. By providing detailed maps that show precisely where these deaths are estimated to have occurred, we provide an important evidence base for looking both to the past, for examples of success, and towards the future, in order to identify where precision public-health initiatives could save the most lives.

The epidemiological toll of child mortality should be considered both in terms of total deaths and as rates of mortality. Focusing only on mortality rates can effectively mask areas in which rates are comparatively low but child deaths are high owing to large population sizes. The number of deaths that occur in high-risk areas has declined, and most under-5 deaths in recent years have occurred in lower-risk areas. This 'prevention paradox'[28] could indicate that whole-population interventions could have a larger overall impact than targeting high-risk areas[29]. At the same time, strategies that target resources to those locations that have the highest number of child deaths risk leaving behind some of the world's most marginalized communities: remote, more-sparsely populated places in which, relative to the number of children born each year, a large number of children die before their fifth birthday. Instead, by considering subnational measures of both counts and rates of deaths of children under 5, decision-makers can better tailor child health programs to align with local contexts, norms and needs. Rural communities with high rates but low counts may benefit from 'last-mile' initiatives to provide effective health services to populations who lack adequate access to care. By contrast, locations with low rates but high counts may require programs that focus on alleviating the cost of care, unsafe environmental exposures or health risks that are uniquely associated with urban slums[30]. The SDGs have pointed the global development agenda towards progress in child survival. Our analysis indicates that reaching the SDG 3.2 targets of 25 child deaths per 1,000 live births and 12 neonatal deaths per 1,000 live births will require only modest improvements or have already been achieved by some units; however, these targets are ambitious for other units in which child mortality remains high. It is worth noting that many countries contain areas that fit both of these profiles. For example, 11 countries

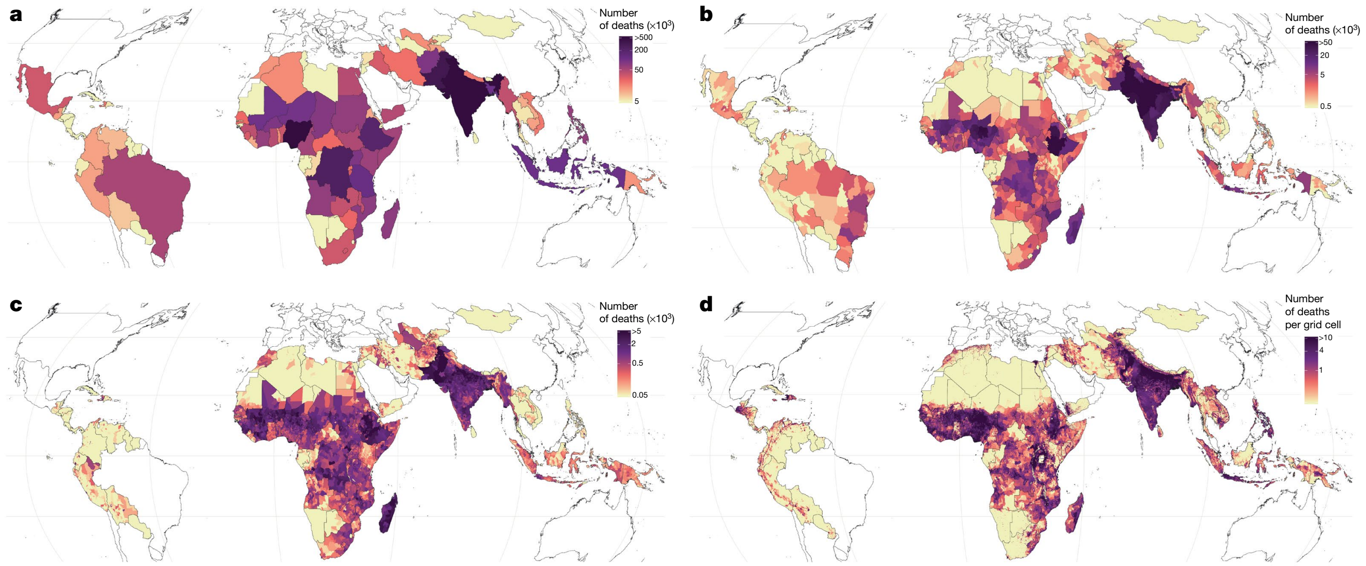


**Fig. 3 | Estimated number of children under 5 who died within 99 countries in 2017. a**, Number of deaths of children under 5 in each country. **b**, Number of deaths in each first administrative-level unit. **c**, Number of deaths in each second administrative-level unit. **d**, Number of deaths of children under 5 in each 5 × 5-km grid cell. Note that scales vary for each aggregation unit.

had at least 1 unit that had already met SDG 3.2 with high certainty, and at least 1 unit that had not. Subnational estimates can empower countries to benchmark gains in child survival against their own subnational exemplars as well as advances that have been achieved by their peers. Through our counterfactual analysis we showed that even if all units had met the SDG 3.2 goal in 2017, there would still have been 2.4 million deaths of children under 5, indicating that 'ending preventable child deaths' is more complex than simply meeting a target threshold. Future research efforts must address the causes of child mortality in local areas and more precisely identify causes of child deaths that are amenable to intervention. To that end, new and innovative data-collection efforts, such as the ongoing Child Health and Mortality Prevention Surveillance network, offer promising prospects by applying high-validity, pathology-based methods alongside verbal autopsies to determine the cause of death[31].

This study offers a unique platform to support the identification of local success stories that could be replicated elsewhere. In Rwanda, for example, the highest U5MR at the district level in 2017 was 60.2% (52.0–67.8%) lower than the lowest U5MR at the district level in 2000. Such gains have been partially credited to focused investments in the country's poorest populations, expanding the Mutuelles de santé insurance program, and developing a strong workforce of community health workers who provide evidence-based treatment and health promotion[32,33]. Nepal and Cambodia are among the exemplars for considerably decreasing subnational inequalities in child survival since 2000. In an era when narrowing disparities within countries is as important as reducing national-level gaps, these results provide the evidence base to inform best practices and stimulate national conversations about related social determinants.

Neonatal mortality rates have also declined but failed to keep pace with reductions in mortality rates of older children, leading to a higher proportion of deaths of children under 5 occurring within the first four weeks of life: from 37.4% (37.1–37.7) in 2000 to 43.7% (43.1–44.3%) in 2017. This trend is probably related to the increase in scale of routine programs and improved infrastructure (for example, vaccination[34], and water and sanitation[35]) and the introduction of effective interventions to target communicable diseases (for example, malaria control[36] and prevention of mother-to-child transmission of HIV[37]). These interventions have tended to target amenable causes of mortality that are more common in older children under 5 rather than dominant causes of neonatal mortality, such as prematurity and congenital anomalies[38]. Notably, irrespective of income level or location, some causes of neonatal death (for example, chromosomal anomalies and severe preterm birth complications) remain difficult to prevent completely with current medical technologies. Ultimately, large gains in neonatal mortality will require serious investment in health system strengthening[39]. Affordable approaches to preventing the majority of neonatal deaths in LMICs exist and there are success stories with lessons learned to apply[40–44], but decisions about which approaches to take must be based on the local epidemiological and health system context. In the absence of spatially detailed cause of death data, subnational neonatal mortality estimates can indicate dominant causes and thus serve as a useful proxy to guide prioritization of interventions[45].

The accuracy and precision of our estimates were primarily determined by the timeliness, quantity and quality of available data. In Sri Lanka, for example, there were no available surveys, and the wide uncertainty intervals surrounding estimates reflect the dearth of available evidence in that country (Extended Data Figs. 3, 4). In certain areas, this decreased the confidence that we had in claiming that a specific subnational area met the SDG 3.2 target (Fig. 1c). This issue is most concerning in cases in which estimated mortality rates are high, thus helping to identify locations in which it would be most useful to focus future data-collection efforts. High mortality rates with large uncertainty intervals were estimated across much of Eastern and Central sub-Saharan Africa, and in Cambodia, Laos, Myanmar and Papua New Guinea (Extended Data Figs. 3, 4). Furthermore, ongoing conflict in countries such as Syria, Yemen and Iraq pose substantial challenges to collecting more contemporaneous data, and our estimates may not fully capture the effects of prolonged civil unrest or war[46,47]. Further methodological and data limitations are discussed in the Methods.

The accurate estimation of mortality is also a matter of equity; highly refined health surveillance is common in high-income countries, whereas in LMICs, in which rates of child mortality are the highest, surveillance that helps to guide investments in health towards the areas with the greatest need is less routine[48]. Ideally, all countries would have high-quality, continuous, and complete civil and vital registration systems that capture all of the births, deaths and causes of death at the appropriate geographical resolution[49]. In the meantime, analyses such as this serve to bridge the information gap that exists between low-mortality countries with strong information systems and countries that face a dual challenge of weaker information systems and higher disease burden.

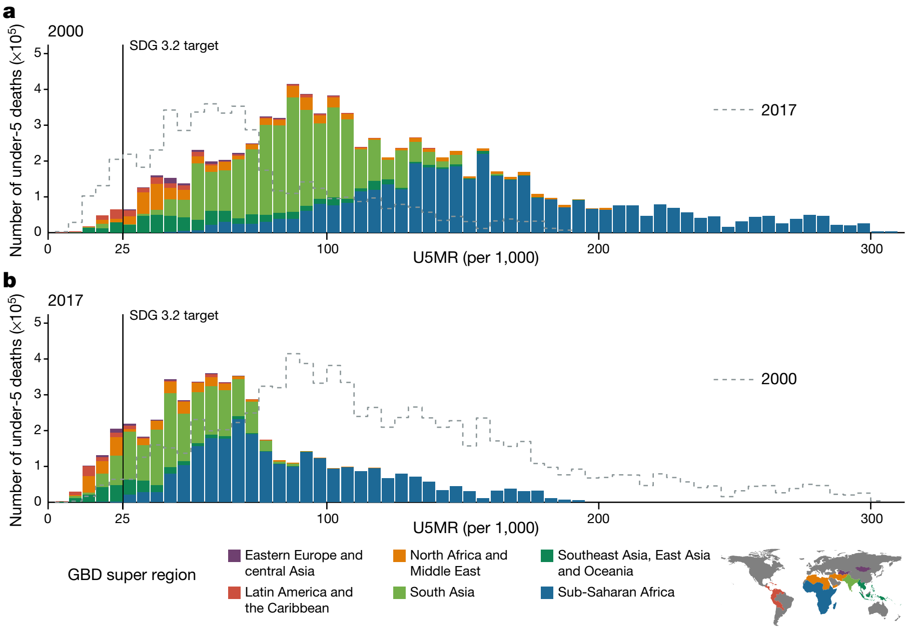


**Fig. 4 | Number of deaths of children under 5, distributed across level of U5MR, in 2000 and in 2017, across 99 countries.** Bar heights represent the total number of deaths of children under 5 within all second administrative-level units with corresponding U5MR. Bins are a width of 5 deaths per 1,000 live births. The colour of each bar represents the global region as defined by the subset legend map. As such, the sum of heights of all bars represents the total number of deaths across the 99 countries. **a**, Deaths of children under 5 in 2000. **b**, Deaths of children under 5 in 2017. The dotted line in the 2000 plot is the shape of the distribution in 2017, and the dotted line in the 2017 plot represents the distribution in 2000.

By harnessing the unprecedented availability of geo-referenced data and developing robust statistical methods, we provide a high-resolution atlas of child death counts and rates since 2000, covering countries that account for 93% of child deaths. We bring attention to subnational geographical inequalities in the distribution, rates and absolute counts of child deaths by age. These high-resolution estimates can help decision-makers to structure policy and program implementation and facilitate pathways to end preventable child deaths[50] by 2030.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1545-0.

Received: 28 March 2019; Accepted: 6 August 2019;
Published online 16 October 2019.

1. Reidpath, D. D. & Allotey, P. Infant mortality rate as an indicator of population health. *J. Epidemiol. Community Health* **57**, 344–346 (2003).
2. United Nations General Assembly. *United Nations Millennium Declaration: Resolution adopted by the General Assembly. A/RES/55/2* (UN General Assembly, 2000).
3. Centers for Disease Control and Prevention. *Ten Great Public Health Achievements—Worldwide, 2001–2010*. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6024a4.htm (2011).
4. GBD 2017 Mortality Collaborators. Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1684–1735 (2018).
5. Golding, N. et al. Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *Lancet* **390**, 2171–2182 (2017).
6. Pezzulo, C. et al. *Geospatial Modeling of Child Mortality across 27 Countries in sub-Saharan Africa. DHS Spatial Analysis Reports No. 13* (USAID, 2016).
7. Li, Z. et al. Changes in the spatial distribution of the under-five mortality rate: small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa. *PLoS ONE* **14**, e0210645 (2019).
8. Lawn, J. E., Cousens, S. & Zupan, J. 4 million neonatal deaths: When? Where? Why? *Lancet* **365**, 891–900 (2005).
9. World Health Organization. SDG 3: Ensure healthy lives and promote wellbeing for all at all ages. https://www.who.int/sdg/targets/en/ (2019).
10. Dwyer-Lindgren, L. et al. Estimation of district-level under-5 mortality in Zambia using birth history data, 1980–2010. *Spat. SpatioTemporal Epidemiol.* **11**, 89–107 (2014).
11. Dwyer-Lindgren, L. et al. Small area estimation of under-5 mortality in Bangladesh, Cameroon, Chad, Mozambique, Uganda, and Zambia using spatially misaligned data. *Popul. Health Metr.* **16**, 13 (2018).
12. Wakefield, J. et al. Estimating under-five mortality in space and time in a developing world context. *Stat. Methods Med. Res.* https://doi.org/10.1177/0962280218767988 (2018).
13. Burke, M., Heft-Neal, S. & Bendavid, E. Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. *Lancet Glob. Health* **4**, e936–e945 (2016).
14. Macharia, P. M. et al. Sub national variation and inequalities in under-five mortality in Kenya since 1965. *BMC Public Health* **19**, 146 (2019).
15. Sousa, A., Hill, K. & Dal Poz, M. R. Sub-national assessment of inequality trends in neonatal and child mortality in Brazil. *Int. J. Equity Health* **9**, 21 (2010).
16. Mohammadi, Y. et al. Measuring Iran's success in achieving Millennium Development Goal 4: a systematic analysis of under-5 mortality at national and subnational levels from 1990 to 2015. *Lancet Glob. Health* **5**, e537–e544 (2017).
17. Wang, Y. et al. Under-5 mortality in 2851 Chinese counties, 1996–2012: a subnational assessment of achieving MDG 4 goals in China. *Lancet* **387**, 273–283 (2016).
18. Horton, R. Offline: in defence of precision public health. *Lancet* **392**, 1504 (2018).
19. Takahashi, S., Metcalf, C. J. E., Ferrari, M. J., Tatem, A. J. & Lessler, J. The geography of measles vaccination in the African Great Lakes region. *Nat. Commun.* **8**, 15585 (2017).
20. Utazi, C. E. et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. *Vaccine* **36**, 1583–1591 (2018).
21. Gething, P. W. et al. Mapping *Plasmodium falciparum* mortality in Africa between 1990 and 2015. *N. Engl. J. Med.* **375**, 2435–2445 (2016).
22. Reiner, R. C. Jr et al. Variation in childhood diarrheal morbidity and mortality in Africa, 2000–2015. *N. Engl. J. Med.* **379**, 1128–1138 (2018).
23. Osgood-Zimmerman, A. et al. Mapping child growth failure in Africa between 2000 and 2015. *Nature* **555**, 41–47 (2018).
24. Amoah, B., Giorgi, E., Heyes, D. J., van Burren, S. & Diggle, P. J. Geostatistical modelling of the association between malaria and child growth in Africa. *Int. J. Health Geogr.* **17**, 7 (2018).
25. Bishai, D. M. et al. Factors contributing to maternal and child mortality reductions in 146 low- and middle-income countries between 1990 and 2010. *PLoS ONE* **11**, e0144908 (2016).
26. Marmot, M. Social determinants of health inequalities. *Lancet* **365**, 1099–1104 (2005).
27. Victora, C. G. et al. Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. *Lancet* **387**, 2049–2059 (2016).
28. Rose, G. Strategy of prevention: lessons from cardiovascular disease. *Br. Med. J. (Clin. Res. Ed.)* **282**, 1847–1851 (1981).
29. Mackenbach, J. P., Lingsma, H. F., van Ravesteyn, N. T. & Kamphuis, C. B. M. The population and high-risk approaches to prevention: quantitative estimates of their contribution to population health in the Netherlands, 1970–2010. *Eur. J. Public Health* **23**, 909–915 (2013).

30. Agarwal, S. & Taneja, S. All slums are not equal: child health conditions among the urban poor. *Indian Pediatr*. **42**, 233–244 (2005).
31. Farag, T. H. et al. Precisely tracking childhood death. *Am. J. Trop. Med. Hyg*. **97**, 3–5 (2017).
32. Farmer, P. E. et al. Reduced premature mortality in Rwanda: lessons from success. *Br. Med. J*. **346**, f65 (2013).
33. Gurusamy, P. S. R. & Janagaraj, P. D. A success story: the burden of maternal, neonatal and childhood mortality in Rwanda - critical appraisal of interventions and recommendations for the future. *Afr. J. Reprod. Health* **22**, 9–16 (2018).
34. McGovern, M. E. & Canning, D. Vaccination and all-cause child mortality from 1985 to 2011: global evidence from the Demographic and Health Surveys. *Am. J. Epidemiol*. **182**, 791–798 (2015).
35. Cheng, J. J., Schuster-Wallace, C. J., Watt, S., Newbold, B. K. & Mente, A. An ecological quantification of the relationships between water, sanitation and infant, child, and maternal mortality. *Environ. Health* **11**, 4 (2012).
36. Steketee, R. W. & Campbell, C. C. Impact of national malaria control scale-up programmes in Africa: magnitude and attribution of effects. *Malar. J*. **9**, 299 (2010).
37. Kiragu, K., Collins, L., Von Zinkernagel, D. & Mushavi, A. Integrating PMTCT into maternal, newborn, and child health and related services: experiences from the global plan priority countries. *J. Acquir. Immune Defic. Syndr*. **75**, S36–S42 (2017).
38. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1736–1788 (2018).
39. Pasha, O. et al. A combined community- and facility-based approach to improve pregnancy outcomes in low-resource settings: a Global Network cluster randomized trial. *BMC Med*. **11**, 215 (2013).
40. Horton, S. et al. Ranking 93 health interventions for low- and middle-income countries by cost-effectiveness. *PLoS ONE* **12**, e0182951 (2017).
41. Simmons, L. E., Rubens, C. E., Darmstadt, G. L. & Gravett, M. G. Preventing preterm birth and neonatal mortality: exploring the epidemiology, causes, and interventions. *Semin. Perinatol*. **34**, 408–415 (2010).
42. Darmstadt, G. L. et al. Evidence-based, cost-effective interventions: how many newborn babies can we save? *Lancet* **365**, 977–988 (2005).
43. Saugstad, O. D. Reducing global neonatal mortality is possible. *Neonatology* **99**, 250–257 (2011).
44. Ntigurirwa, P. et al. A health partnership to reduce neonatal mortality in four hospitals in Rwanda. *Glob. Health* **13**, 28 (2017).
45. Knippenberg, R. et al. Systematic scaling up of neonatal care in countries. *Lancet* **365**, 1087–1098 (2005).
46. GBD 2015 Eastern Mediterranean Region Neonatal, Infant, and under-5 Mortality Collaborators. Neonatal, infant, and under-5 mortality and morbidity burden in the Eastern Mediterranean region: findings from the Global Burden of Disease 2015 study. *Int. J. Public Health* **63**, 63–77 (2018).
47. Wagner, Z. et al. Armed conflict and child mortality in Africa: a geospatial analysis. *Lancet* **392**, 857–865 (2018).
48. Mikkelsen, L. et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet* **386**, 1395–1406 (2015).
49. AbouZahr, C. et al. Civil registration and vital statistics: progress in the data revolution for counting and accountability. *Lancet* **386**, 1373–1385 (2015).
50. Annan, K. Data can help to end malnutrition across Africa. *Nature* **555**, 7 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Roy Burstein[1,612], Nathaniel J. Henry[1,612], Michael L. Collison[1], Laurie B. Marczak[1], Amber Sligar[1], Stefanie Watson[1], Neal Marquez[1], Mahdieh Abbasalizad-Farhangi[2], Masoumeh Abbasi[3], Foad Abd-Allah[4], Amir Abdoli[5], Mohammad Abdollahi[6], Ibrahim Abdollahpour[7,8], Rizwan Suliankatchi Abdulkader[9], Michael R. M. Abrigo[10], Dilaram Acharya[11,12], Oladimeji M. Adebayo[13], Victor Adekanmbi[14], Davoud Adham[15], Mahdi Afshari[16], Mohammad Aghaali[17], Keivan Ahmadi[18], Mehdi Ahmadi[19], Ehsan Ahmadpour[20], Rushdia Ahmed[21,22], Chalachew Genet Akal[23], Joshua O. Akinyemi[24], Fares Alahdab[25], Noore Alam[26], Genet Melak Alamene[27], Kefyalew Addis Alene[28,29], Mehran Alijanzadeh[30], Cyrus Alinia[31], Vahid Alipour[32,33], Syed Mohamed Aljunid[34,35], Mohammed J. Almalki[36,37], Hesham M. Al-Mekhlafi[38,39], Khalid Altirkawi[40], Nelson Alvis-Guzman[41,42], Adeladza Kofi Amegah[43], Saeed Amini[44], Arianna Maever Loreche Amit[45,46], Zohreh Anbari[44], Sofia Androudi[47], Mina Anjomshoa[48], Fereshteh Ansari[49], Carl Abelardo T. Antonio[50,51], Jalal Arabloo[33], Zohreh Arefi[52], Olatunde Aremu[53], Bahram Armoon[54,55], Amit Arora[56,57], Al Artaman[58], Anvar Asadi[59], Mehran Asadi-Aliabadi[60], Amir Ashraf-Ganjouei[7,61], Reza Assadi[62], Bahar Ataeinia[63], Sachin R. Atre[64,65], Beatriz Paulina Ayala Quintanilla[66,67], Martin Amogre Ayanore[68], Samad Azari[33], Ebrahim Babaee[60], Arefeh Babazadeh[69], Alaa Badawi[70,71], Soghra Bagheri[3], Mojtaba Bagherzadeh[72], Nafiseh Baheiraei[73,74], Abbas Balouchi[75], Aleksandra Barac[76,77], Quique Bassat[78,79], Bernhard T. Baune[80], Mohsen Bayati[81], Neeraj Bedi[37,82], Ettore Beghi[83], Masoud Behzadifar[84], Meysam Behzadifar[85], Yared Belete Belay[86], Brent Bell[1], Michelle L. Bell[87], Dessalegn Ajema Berbada[88], Robert S. Bernstein[89,90], Natalia V. Bhattacharjee[1], Suraj Bhattarai[91,92], Zulfiqar A. Bhutta[93,94], Ali Bijani[95], Somayeh Bohlouli[96], Nicholas J. K. Breitborde[97,98], Gabrielle Britton[99,100], Annie J. Browne[101], Sharath Burugina Nagaraja[102], Reinhard Busse[103], Zahid A. Butt[104,105], Josip Car[106,107], Rosario Cárdenas[108], Carlos A. Castañeda-Orjuela[109,110], Ester Cerin[111,112], Wagaye Fentahun Chanie[113], Pranab Chatterjee[114], Dinh-Toi Chu[115], Cyrus Cooper[116,117], Vera M. Costa[118], Koustuv Dalal[119,120], Lalit Dandona[1,121], Rakhi Dandona[1,121], Farah Daoud[1], Ahmad Daryani[122], Rajat Das Gupta[21], Ian Davis[1], Nicole Davis Weaver[1], Dragos Virgil Davitoiu[123,124], Jan-Walter De Neve[125], Feleke Mekonnen Demeke[126], Gebre Teklemariam Demoz[127,128], Kebede Deribe[129,130], Rupak Desai[131], Aniruddha Deshpande[1], Hanna Demelash Desyibelew[132], Sagnik Dey[133], Samath Dhamminda Dharmaratne[1,134], Meghnath Dhimal[135], Daniel Diaz[136,137], Leila Doshmangir[138], Andre R. Duraes[139,140], Laura Dwyer-Lindgren[1,141], Lucas Earl[1], Roya Ebrahimi[142], Soheil Ebrahimpour[69], Andem Effiong[143], Aziz Eftekhari[144,145], Elham Ehsani-Chimeh[146], Iman El Sayed[147], Maysaa El Sayed Zaki[148], Maha El Tantawi[149,150], Ziad El-Khatib[151], Mohammad Hassan Emamian[152], Shymaa Enany[153], Sharareh Eskandarieh[7], Oghenowede Eyawo[154,155], Maha Ezalarab[1], Mahbobeh Faramarzi[156], Mohammad Fareed[157], Roghiyeh Faridnia[158], Andre Faro[159], Ali Akbar Fazaeli[160], Mehdi Fazlzadeh[161,162], Netsanet Fentahun[163], Seyed-Mohammad Fereshtehnejad[164,165], João C. Fernandes[166], Irina Filip[167,168], Florian Fischer[169], Nataliya A. Foigt[170], Masoud Foroutan[171], Joel Msafiri Francis[172], Takeshi Fukumoto[173,174], Nancy Fullman[1], Silvano Gallus[175], Destallem Gebremedhin Gebre[176,177], Tsegaye Tewelde Gebrehiwot[178], Gebreamlak Gebremedhn Gebremeskel[179,180], Bradford D. Gessner[181,182], Birhanu Geta[183], Peter W. Gething[101], Reza Ghadimi[184], Keyghobad Ghadiri[3], Mahsa Ghajarzadeh[185], Ahmad Ghashghaee[186], Paramjit Singh Gill[187], Tiffany K. Gill[188], Nick Golding[609], Nelson G. M. Gomes[189,190], Philimon N. Gona[191], Sameer Vali Gopalani[192,193], Giuseppe Gorini[194], Bárbara Niegia Garcia Goulart[195], Nicholas Graetz[1], Felix Greaves[196,197], Manfred S. Green[198], Yuming Guo[199,200], Arvin Haj-Mirzaian[201,202], Arya Haj-Mirzaian[201,203], Brian James Hall[204], Samer Hamidi[205], Hamidreza Haririan[206], Josep Maria Haro[207,208], Milad Hasankhani[209], Edris Hasanpoor[210], Amir Hasanzadeh[211,212], Hadi Hassankhani[213,214], Hamid Yimam Hassen[215,216], Mohamed I. Hegazy[4], Delia Hendrie[217], Fatemeh Heydarpour[218], Thomas R. Hird[199,219], Chi Linh Hoang[220], Gillian Hollerich[1], Enayatollah Homaie Rad[221,222], Mojtaba Hoseini-Ghahfarokhi[223], Naznin Hossain[22,224], Mostafa Hosseini[225], Mehdi Hosseinzadeh[33,226], Mihaela Hostiuc[123,227], Sorin Hostiuc[228,229], Mowafa Househ[230,231], Mohamed Hsairi[232], Olayinka Stephen Ilesanmi[233], Mohammad Hasan Imani-Nasab[234], Usman Iqbal[235], Seyed Sina Naghibi Irvani[236], Nazrul Islam[237,238], Sheikh Mohammed Shariful Islam[239,240], Mikk Jürisson[241], Nader Jafari Balalami[242], Amir Jalali[243], Javad Javidnia[244], Achala Upendra Jayatilleke[245,246], Ensiyeh Jenabi[247], John S. Ji[248], Yash B. Jobanputra[249], Kimberly Johnson[1], Jost B. Jonas[250,251], Zahra Jorjoran Shushtari[252], Jacek Jerzy Jozwiak[253,254], Ali Kabir[255], Amaha Kahsay[256], Hamed Kalani[257,258], Rohollah Kalhor[259], Manoochehr Karami[260], Surendra Karki[261,262], Amir Kasaeian[263,264], Nicholas J. Kassebaum[1,265], Peter Njenga Keiyoro[266], Grant Rodgers Kemp[1,267], Roghayeh Khabiri[268,269], Yousef Saleh Khader[270], Morteza Abdullatif Khafaie[271], Ejaz Ahmad Khan[272], Junaid Khan[273], Muhammad Shahzeb Khan[274,275], Young-Ho Khang[276,277], Khaled Khatab[278,279], Amir Khater[280], Mona M. Khater[281], Alireza Khatony[3], Mohammad Khazaei[282], Salman Khazaei[260], Maryam Khazaei-Pool[283], Jagdish Khubchandani[284], Neda Kianipour[3,285], Yun Jin Kim[286], Ruth W. Kimokoti[287], Damaris K. Kinyoki[1,141], Adnan Kisa[288,289], Sezer Kisa[290], Tufa Kolola[291], Soewarta Kosen[292], Parvaiz A. Koul[293], Ai Koyanagi[79,294], Moritz U. G. Kraemer[295,296], Kewal Krishan[297], Kris J. Krohn[1], Nuworza Kugbey[298,299], G. Anil Kumar[121], Manasi Kumar[300,301], Pushpendra Kumar[302], Desmond Kuupiel[303,304], Ben Lacey[305,306], Sheetal D. Lad[307], Faris Hasan Lami[308], Anders O. Larsson[309,310], Paul H. Lee[311], Mostafa Leili[282], Aubrey J. Levine[1], Shanshan Li[199], Lee-Ling Lim[312,313], Stefan Listl[314,315], Joshua Longbottom[316], Jaifred Christian F. Lopez[45,317], Stefan Lorkowski[318,319], Sameh Magdeldin[320,321], Hassan Magdy Abd El Razek[322], Muhammed Magdy Abd El Razek[323], Azeem Majeed[196], Afshin Maleki[142], Reza Malekzadeh[324,325], Deborah Carvalho Malta[326], Abdullah A. Mamun[327], Navid Manafi[328,329], Ana-Laura Manda[330], Morteza Mansourian[331], Francisco Rogerlândio Martins-Melo[332], Anthony Masaka[333], Benjamin Ballard Massenburg[334], Pallab K. Maulik[335,336], Benjamin K. Mayala[1], Mohsen Mazidi[337], Martin McKee[338], Ravi Mehrotra[339], Kala M. Mehta[340], Gebrekiros Gebremichael Meles[88], Walter Mendoza[341], Ritesh G. Menezes[342], Atte Meretoja[343,344], Tuomo J. Meretoja[345,346], Tomislav Mestrovic[347,348], Ted R. Miller[217,349], Molly K. Miller-Petrie[1], Edward J. Mills[350], George J. Milne[351], G. K. Mini[352], Seyed Mostafa Mir[353,354], Hamed Mirjalali[355], Erkin M. Mirrakhimov[356,357], Efat Mohamadi[358], Dara K. Mohammad[359,360], Aso Mohammad Darwesh[361], Naser Mohammad Gholi Mezerji[362], Ammas Siraj Mohammed[363], Shafiu Mohammed[364,365], Ali H. Mokdad[1,141], Mariam Molokhia[366], Lorenzo Monasta[367], Yoshan Moodley[303], Mahmood Moosazadeh[368], Ghobad Moradi[369,370], Masoud Moradi[3,59], Yousef Moradi[371], Maziar Moradi-Lakeh[60], Mehdi Moradinazar[59], Paula Moraga[372], Lidia Morawska[373], Abbas Mosapour[353,374], Seyyed Meysam Mousavi[375], Ulrich Otto Mueller[376,377], Atalay Goshu Muluneh[378], Ghulam Mustafa[379,380], Behnam Nabavizadeh[381], Mehdi Naderi[382], Ahamarshan Jayaraman Nagarajan[383,384], Azin Nahvijou[385], Farid Najafi[386], Vinay Nangia[387], Duduzile Edith Ndwandwe[388], Nahid Neamati[353], Ionut Negoi[389,390], Ruxandra Irina Negoi[391,392], Josephine W. Ngunjiri[393], Huong Lan Thi Nguyen[394], Long Hoang Nguyen[395], Son Hoang Nguyen[395], Katie R. Nielsen[396,397], Dina Nur Anggraini Ningrum[610,611], Yirga Legesse Nirayo[398], Molly R. Nixon[1], Chukwudi A. Nnaji[388,399], Marzieh Nojomi[60,400], Mehdi Noroozi[401], Shirin Nosratnejad[402], Jean Jacques Noubiap[403], Soraya Nouraei Motlagh[234], Richard Ofori-Asenso[404,405], Felix Akpojene Ogbo[406], Kelechi E. Oladimeji[303,407], Andrew T. Olagunju[408,409], Meysam Olfatifar[410], Solomon Olum[411,412], Bolajoko Olubukunola Olusanya[413], Mojisola Morenike Oluwasanu[414], Obinna E. Onwujekwe[415], Eyal Oren[416,417], Doris D. V. Ortega-Altamirano[418], Alberto Ortiz[419,420], Osayomwanbo Osarenotor[421], Frank B. Osei[422,423], Aaron E. Osgood-Zimmerman[1], Stanislav S. Otstavnov[424,425], Mayowa Ojo Owolabi[426], Mahesh P. A.[427], Abdol Sattar Pagheh[122], Smita Pakhale[428], Songhomitra Panda-Jonas[429], Animika Pandey[121], Eun-Kee Park[430], Hadi Parsian[353], Tahereh Pashaei[142], Sangram Kishor Patel[431,432], Veincent Christian Filipino Pepito[433], Alexandre Pereira[434,435], Samantha Perkins[1], Brandon V. Pickering[1], Thomas Pilgrim[436], Majid Pirestani[437], Bakhtiar Piroozi[370], Meghdad Pirsaheb[3], Oleguer Plana-Ripoll[438], Hadi Pourjafar[439,440], Parul Puri[302], Mostafa Qorbani[441], Hedley Quintana[100], Mohammad Rabiee[442], Navid Rabiee[72], Amir Radfar[443,444], Alireza Rafiei[445,446], Fakher Rahim[447,448], Zohreh Rahimi[449], Vafa Rahimi-Movaghar[450], Shadi Rahimzadeh[63], Fatemeh Rajati[451], Sree Bhushan Raju[452], Azra Ramezankhani[453,454], Chhabi Lal Ranabhat[455,456], Davide Rasella[457,458], Vahid Rashedi[459], Lal Rawal[56,460], Robert C. Reiner Jr[1,141], Andre M. N. Renzaho[461], Satar Rezaei[462], Aziz Rezapour[33], Seyed Mohammad Riahi[463,464], Ana Isabel Ribeiro[465], Leonardo Roever[466], Elias Merdassa Roro[467,468], Max Roser[469], Gholamreza Roshandel[324,470], Daem Roshani[471], Ali Rostami[472], Enrico Rubagotti[473,474], Salvatore Rubino[475], Siamak Sabour[463], Nafis Sadat[1], Ehsan Sadeghi[59], Reza Saeedi[476], Yahya Safari[3], Roya Safari-Faramani[477], Mahdi Safdarian[450,478], Amirhossein Sahebkar[479,480], Mohammad Reza Salahshoor[481], Nasir Salam[482], Payman Salamati[450,483], Farkhonde Salehi[484], Saleh Salehi Zahabi[485,486], Yahya Salimi[386], Hamideh Salimzadeh[324], Joshua A. Salomon[487], Evanson Zondani Sambala[388], Abdallah M. Samy[488], Milena M. Santric Milicevic[489], Bruno Piassi Sao Jose[490], Sivan Yegnanarayana Iyer Saraswathy[491,492], Rodrigo Sarmiento-Suárez[493], Benn Sartorius[141,494], Brijesh Sathian[495,496], Sonia Saxena[497], Alyssa N. Sbarra[1], Lauren E. Schaeffer[1], David C. Schwebel[498], Sadaf G. Sepanlou[324,325], Seyedmojtaba Seyedmousavi[499,500], Faramarz Shaahmadi[501], Masood Ali Shaikh[502], Mehran Shams-Beyranvand[63,503], Amir Shamshirian[504], Morteza Shamsizadeh[505], Kiomars Sharafi[3], Mehdi Sharif[506,507], Mahdi Sharif-Alhoseini[450], Hamid Sharifi[508], Jayendra Sharma[509], Rajesh Sharma[510], Aziz Sheikh[511,512], Chloe Shields[1], Mika Shigematsu[513], Rahman Shiri[514], Ivy Shiue[515], Kerem Shuval[198], Tariq J. Siddiqi[275], João Pedro Silva[118], Jasvinder A. Singh[516,517], Dhirendra Narain Sinha[518,519], Malede Mequanent Sisay[378,520], Solomon Sisay[521], Karen Sliwa[403], David L. Smith[1,141], Ranjani Somayaji[522,523], Moslem Soofi[524], Joan B. Soriano[525,526], Chandrashekhar T. Sreeramareddy[527], Agus Sudaryanto[528], Mu'awiyyah Babale Sufiyan[529], Bryan L. Sykes[530], P. N. Sylaja[531], Rafael Tabarés-Seisdedos[532,533], Karen M. Tabb[534], Takahiro Tabuchi[535], Nuno Taveira[536,537], Mohamad-Hani Temsah[538,539], Abdullah Sulieman Terkawi[540,541], Zemenu Tadesse Tessema[378], Kavumpurathu Raman Thankappan[542], Sathish Thirunavukkarasu[543], Quyen G. To[544], Marcos Roberto Tovani-Palone[545], Bach Xuan Tran[546], Khanh Bao Tran[547,548], Irfan Ullah[549,550], Muhammad Shariq Usman[275], Olalekan A. Uthman[551], Amir Vahedian-Azimi[552,553], Pascual R. Valdez[554,555], Job F. M. van Boven[556,557], Tommi Juhani Vasankari[558], Yasser Vasseghian[462], Yousef Veisani[559], Narayanaswamy Venketasubramanian[560,561], Francesco S. Violante[562,563], Sergey Konstantinovitch Vladimirov[564,565], Vasily Vlassov[566], Theo Vos[1,141], Giang Thu Vu[220], Isidora S. Vujcic[77], Yasir Waheed[567], Jon Wakefield[568], Haidong Wang[1,141], Yafeng Wang[569], Yuan-Pang Wang[570], Joseph L. Ward[571], Robert G. Weintraub[572,573], Kidu Gidey Weldegwergs[398], Girmay Teklay Weldesamuel[574], Ronny Westerman[575], Charles Shey Wiysonge[576,577], Dawit Zewdu Wondafrash[578,579], Lauren Woyczynski[1], Ai-Min Wu[580], Gelin Xu[581], Abbas Yadegar[355], Tomohide Yamada[582], Vahid Yazdi-Feyzabadi[583,584], Christopher Sabo Yilgwan[585,586], Paul Yip[587,588], Naohiro Yonemoto[589], Javad Yoosefi Lebni[331], Mustafa Z. Younis[590,591], Mahmoud Yousefifard[592], Hebat-Allah Salah A. Yousof[593], Chuanhua Yu[569,594], Hasan Yusefzadeh[595], Erfan Zabeh[596,597], Telma Zahirian Moghadam[33,598], Sojib Bin Zaman[599,600], Mohammad Zamani[601], Hamed Zandian[598,602], Alireza Zangeneh[603], Taddese Alemu Zerfu[604,605], Yunquan Zhang[606,607], Arash Ziapour[331], Sanjay Zodpey[608], Christopher J. L. Murray[1,141,613] & Simon I. Hay[1,141,613]*

[1]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA. [2]Department of Nutrition, Tabriz University of Medical Sciences, Tabriz, Iran. [3]Kermanshah University of Medical Sciences, Kermanshah, Iran. [4]Department of Neurology, Cairo University, Cairo, Egypt. [5]Department of Parasitology and Mycology, Jahrom University of Medical Sciences, Jahrom, Iran. [6]The Institute of Pharmaceutical Sciences, Tehran University of Medical Sciences, Tehran, Iran. [7]Multiple Sclerosis Research Center, Tehran University of Medical Sciences, Tehran, Iran. [8]Department of Epidemiology, Arak University of Medical Sciences, Arak, Iran. [9]Department of Public Health, Ministry of Health, Riyadh, Saudi Arabia. [10]Research Department, Philippine Institute for Development Studies, Quezon City, The Philippines. [11]Department of Preventive Medicine, Dongguk University, Gyeongju, South Korea. [12]Department of Community Medicine, Kathmandu University, Devdaha, Nepal. [13]College of Medicine, University College Hospital, Ibadan, Nigeria. [14]School of Medicine, Cardiff University, Cardiff, UK. [15]School of Health, Ardabil University of Medical Science, Ardabil, Iran. [16]Department of Community Medicine, Zabol University of Medical Sciences, Zabol, Iran. [17]Department of Epidemiology and Biostatistics, Qom University of Medical Sciences, Qom, Iran. [18]School of Pharmacy, University of Lincoln, Lincoln, UK. [19]Environmental Technologies Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. [20]Department of Parasitology and Mycology, Tabriz University of Medical Sciences, Tabriz, Iran. [21]James P. Grant School of Public Health, Brac University, Dhaka, Bangladesh. [22]Health Systems and Population Studies Division, International Centre for Diarrhoeal Disease Research Bangladesh, Dhaka, Bangladesh. [23]Department of Medical Laboratory Science, Bahir Dar University, Bahir

Dar, Ethiopia. [24]Epidemiology and Medical Statistics, University of Ibadan, Ibadan, Nigeria. [25]Evidence Based Practice Center, Mayo Clinic Foundation for Medical Education and Research, Rochester, MN, USA. [26]Prevention Division, Queensland Health, Herston, Queensland, Australia. [27]School of Health Sciences, Madda Walabu University, Bale Goba, Ethiopia. [28]Institute of Public Health, University of Gondar, Gondar, Ethiopia. [29]Research School of Population Health, Australian National University, Canberra, Australian Capital Territory, Australia. [30]Qazvin University of Medical Sciences, Qazvin, Iran. [31]Department of Health Care Management and Economics, Urmia University of Medical Science, Urmia, Iran. [32]Health Economics Department, Iran University of Medical Sciences, Tehran, Iran. [33]Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran. [34]Department of Health Policy and Management, Kuwait University, Safat, Kuwait. [35]International Centre for Casemix and Clinical Coding, National University of Malaysia, Bandar Tun Razak, Malaysia. [36]Faculty of Public Health and Tropical Medicine, Jazan University, Jazan, Saudi Arabia. [37]Jazan University, Jazan, Saudi Arabia. [38]Medical Research Center, Jazan University, Jazan, Saudi Arabia. [39]Department of Medical Parasitology, Sana'a University, Sana'a, Yemen. [40]King Saud University, Riyadh, Saudi Arabia. [41]Research Group in Health Economics, Universidad de Cartagena, Cartagena, Colombia. [42]Research Group in Hospital Management and Health Policies, Universidad de la Costa, Barranquilla, Colombia. [43]Biomedical Science, University of Cape Coast, Cape Coast, Ghana. [44]Health Services Management Department, Arak University of Medical Sciences, Arak, Iran. [45]Department of Epidemiology and Biostatistics, University of the Philippines Manila, Manila, The Philippines. [46]Online Programs for Applied Learning, Johns Hopkins University, Baltimore, MD, USA. [47]Department of Medicine, University of Thessaly, Volos, Greece. [48]Social Determinants of Health Research Center, Rafsanjan University of Medical Sciences, Rafsanjan, Iran. [49]Research Center for Evidence Based Medicine-Health Management and Safety Promotion Research Institute, Tabriz University of Medical Sciences, Tabriz, Iran. [50]Department of Health Policy and Administration, University of the Philippines Manila, Manila, The Philippines. [51]Department of Applied Social Sciences, Hong Kong Polytechnic University, Hong Kong, China. [52]Department of Health Promotion and Education, Tehran University of Medical Sciences, Tehran, Iran. [53]School of Health Sciences, Birmingham City University, Birmingham, UK. [54]School of Nursing and Midwifery, Saveh University of Medical Sciences, Saveh, Iran. [55]Social Determinants of Health Research Center, Saveh University of Medical Sciences, Saveh, Iran. [56]School of Science and Health, Western Sydney University, Penrith, New South Wales, Australia. [57]Oral Health Services, Sydney Local Health District, Sydney, New South Wales, Australia. [58]Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. [59]Research Center for Environmental Determinants of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. [60]Preventive Medicine and Public Health Research Center, Iran University of Medical Sciences, Tehran, Iran. [61]Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran. [62]Education Development Center, Mashhad University of Medical Sciences, Mashhad, Iran. [63]Non-communicable Diseases Research Center, Tehran University of Medical Sciences, Tehran, Iran. [64]Center for Clinical Global Health Education, Johns Hopkins University, Baltimore, MD, USA. [65]Dr D. Y. Patil Medical College, Pune, India. [66]The Judith Lumley Centre, La Trobe University, Melbourne, Victoria, Australia. [67]General Office for Research and Technological Transfer, Peruvian National Institute of Health, Lima, Peru. [68]Department of Family and Community Health, University of Health and Allied Sciences, Ho, Ghana. [69]Center for Infectious Diseases Research, Babol, Iran. [70]Public Health Risk Sciences Division, Public Health Agency of Canada, Toronto, Ontario, Canada. [71]Department of Nutritional Sciences, University of Toronto, Toronto, Ontario, Canada. [72]Department of Chemistry, Sharif University of Technology, Tehran, Iran. [73]Tissue Engineering and Applied Cell Sciences Division, Tarbiat Modares University, Tehran, Iran. [74]Division of Diseases, Advanced Technologies Research Group, Tehran, Iran. [75]School of Nursing and Midwifery, Iran University of Medical Sciences, Tehran, Iran. [76]Clinic for Infectious and Tropical Diseases, Clinical Center of Serbia, Belgrade, Serbia. [77]Faculty of Medicine, University of Belgrade, Belgrade, Serbia. [78]Barcelona Institute for Global Health, University of Barcelona, Barcelona, Spain. [79]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain. [80]Department of Psychiatry, Melbourne Medical School, Melbourne, Victoria, Australia. [81]Health Human Resources Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. [82]Department of Community Medicine, Gandhi Medical College Bhopal, Bhopal, India. [83]Department of Neuroscience, Mario Negri Institute for Pharmacological Research, Milan, Italy. [84]Social Determinants of Health Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran. [85]Hepatitis Research Center, Lorestan University of Medical Sciences, Khorramabad, Iran. [86]Pharmacoepidemiology and Social Pharmacy, Mekelle University, Mekelle, Ethiopia. [87]School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA. [88]Department of Public Health, Arba Minch University, Arba Minch, Ethiopia. [89]Hubert Department of Global Health, Emory University, Atlanta, GA, USA. [90]Department of Global Health, University of South Florida, Tampa, FL, USA. [91]London School of Hygiene & Tropical Medicine, London, UK. [92]Nepal Academy of Science & Technology, Patan, Nepal. [93]The Centre for Global Child Health, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. [94]Center of Excellence in Women and Child Health, Aga Khan University, Karachi, Pakistan. [95]Social Determinants of Health Research Center, Babol University of Medical Sciences, Babol, Iran. [96]Department of Veterinary Medicine, Karaj Islamic Azad University, Kermanshah, Iran. [97]Department of Psychology, Ohio State University, Columbus, OH, USA. [98]Psychiatry and Behavioral Health Department, Ohio State University, Columbus, OH, USA. [99]Neuroscience Department, Institute for Scientific Research and High Technology Services, City of Knowledge, Panama. [100]Gorgas Memorial Institute for Health Studies, Panama, Panama. [101]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [102]Department of Community Medicine, Employees' State Insurance Model Hospital, Bangalore, India. [103]Department for Health Care Management, Technical University of Berlin, Berlin, Germany. [104]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada. [105]Al Shifa School of Public Health, Al Shifa Trust Eye Hospital, Rawalpindi, Pakistan. [106]Centre for Population Health Sciences, Nanyang Technological University, Singapore, Singapore. [107]Global Ehealth Unit, Imperial College London, London, UK. [108]Department of Population and Health, Metropolitan Autonomous University, Mexico City, Mexico. [109]Colombian National Health Observatory, National Institute of Health, Bogota, Colombia. [110]Epidemiology and Public Health Evaluation Group, National University of Colombia, Bogota, Colombia. [111]Mary Mackillop Institute for Health Research, Australian Catholic University, Melbourne, Victoria, Australia. [112]School of Public Health, University of Hong Kong, Hong Kong, China. [113]Department of Obstetrics and Gynaecology, University of Gondar, Gondar, Ethiopia. [114]Division of Epidemiology, National Institute of Cholera and Enteric Diseases, Kolkata, India. [115]Faculty of Biology, Hanoi National University of Education, Hanoi, Vietnam. [116]Department of Rheumatology, University of Oxford, Oxford, UK. [117]Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK. [118]Applied Molecular Biosciences Unit (UCIBIO), University of Porto, Porto, Portugal. [119]Institute of Public Health Kalyani, Kalyani, India. [120]School of Health Science, Orebro University, Orebro, Sweden. [121]Public Health Foundation of India, Gurugram, India. [122]Toxoplasmosis Research Center, Mazandaran University of Medical Sciences, Sari, Iran. [123]Department of General Surgery, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [124]Department of Surgery, Clinical Emergency Hospital St Pantelimon, Bucharest, Romania. [125]Heidelberg Institute of Global Health (HIGH), Heidelberg University, Heidelberg, Germany. [126]Bahir Dar University, Bahir Dar, Ethiopia. [127]School of Pharmacy, Aksum University, Aksum, Ethiopia. [128]Addis Ababa University, Addis Ababa, Ethiopia. [129]School of Public Health, Addis Ababa University, Addis Ababa, Ethiopia. [130]Department of Global Health and Infection, Brighton and Sussex Medical School, Brighton, UK. [131]Division of Cardiology, Atlanta Veterans Affairs Medical Center, Decatur, GA, USA. [132]Public Health Nutrition, Bahir Dar University, Bahir Dar, Ethiopia. [133]Centre for Atmospheric Sciences, Indian Institute of Technology Delhi, New Delhi, India. [134]Department of Community Medicine, University of Peradeniya, Peradeniya, Sri Lanka. [135]Health Research Section, Nepal Health Research Council, Kathmandu, Nepal. [136]Center of Complexity Sciences, National Autonomous University of Mexico, Mexico City, Mexico. [137]Facultad de Medicina Veterinaria y Zootecnia, Autonomous University of Sinaloa, Culiacan Rosales, Mexico. [138]Department of Health Policy and Economy, Tabriz University of Medical Sciences, Tabriz, Iran. [139]School of Medicine, Federal University of Bahia, Salvador, Brazil. [140]Diretoria Médica, Roberto Santos General Hospital, Salvador, Brazil. [141]Department of Health Metrics Sciences, School of Medicine, University of Washington, Seattle, WA, USA. [142]Environmental Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran. [143]Clinical Epidemiology and Biostatistics, University of Newcastle, Newcastle, New South Wales, Australia. [144]Department of Toxicology and Pharmacology, Tabriz University of Medical Sciences, Tabriz, Iran. [145]Department of Basic Sciences, Maragheh University of Medical Sciences, Maragheh, Iran. [146]National Institute for Health Researchers, Tehran University of Medical Sciences, Tehran, Iran. [147]Medical Research Institute, Alexandria University, Alexandria, Egypt. [148]Department of Clinical Pathology, Mansoura University, Mansoura, Egypt. [149]Pediatric Dentistry and Dental Public Health, Alexandria University, Alexandria, Egypt. [150]Preventive Dental Sciences, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. [151]Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden. [152]Ophthalmic Epidemiology Research Center, Shahroud University of Medical Sciences, Shahroud, Iran. [153]Department of Microbiology and Immunology, Suez Canal University, Ismailia, Egypt. [154]Epidemiology and Population Health, British Columbia Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada. [155]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada. [156]Babol University of Medical Sciences, Babol, Iran. [157]College of Medicine, Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. [158]Department of Parasitology, Mazandaran University of Medical Sciences, Sari, Iran. [159]Department of Psychology, Federal University of Sergipe, Sao Cristovao, Brazil. [160]Social Determinants of Health Research Center, Hamadan University of Medical Sciences, Hamadan, Iran. [161]Environmental Health Engineering, Tehran University of Medical Sciences, Tehran, Iran. [162]Department of Environmental Health Engineering, Ardabil University of Medical Science, Ardabil, Iran. [163]Department of Public Health Nutrition, Bahir Dar University, Bahir Dar, Ethiopia. [164]Department of Neurobiology, Karolinska Institutet, Stockholm, Sweden. [165]Division of Neurology, University of Ottawa, Ottawa, Ontario, Canada. [166]Center for Biotechnology and Fine Chemistry, Catholic University of Portugal, Porto, Portugal. [167]Psychiatry Department, Kaiser Permanente, Fontana, CA, USA. [168]Department of Health Sciences, A.T. Still University, Mesa, AZ, USA. [169]Department of Public Health Medicine, Bielefeld University, Bielefeld, Germany. [170]Institute of Gerontology, National Academy of Medical Sciences of Ukraine, Kyiv, Ukraine. [171]Abadan School of Medical Sciences, Abadan, Iran. [172]Clinical Medicine and Wits Reproductive Health and HIV Institute, University of the Witwatersrand, Johannesburg, South Africa. [173]Gene Expression & Regulation Program, Cancer Institute (W.I.A.), Philadelphia, PA, USA. [174]Department of Dermatology, Kobe University, Kobe, Japan. [175]Department of Environmental Health Science, Mario Negri Institute for Pharmacological Research, Milan, Italy. [176]Mekelle University, Mekelle, Ethiopia. [177]Dr Tewelde Legesse Health Sciences College, Mekelle, Ethiopia. [178]Department of Epidemiology, Jimma University, Jimma, Ethiopia. [179]School of Nursing, Mekelle University, Mekelle, Ethiopia. [180]Nursing Department, Aksum University, Aksum, Ethiopia. [181]Vaccines Department, Pfizer, Collegeville, PA, USA. [182]Agency of Preventive Medicine, Paris, France. [183]Department of Pharmacy, Wollo University, Dessie, Ethiopia. [184]Health Research Institute, Babol University of Medical Sciences, Babol, Iran. [185]Department of Neurology, Tehran University of Medical Sciences, Tehran, Iran. [186]Department of Health Services Management, Iran University of Medical Sciences, Tehran, Iran. [187]Unit of Academic Primary Care, University of Warwick, Coventry, UK. [188]Adelaide Medical School, University of Adelaide, Adelaide, South Australia, Australia. [189]Department of Chemistry, University of Porto, Porto, Portugal. [190]REQUIMTE/LAQV, Porto, Portugal. [191]Nursing and Health Sciences Department, University of Massachusetts Boston, Boston, MA, USA. [192]Department of Biostatistics and Epidemiology, University of Oklahoma, Oklahoma City, OK, USA. [193]Department of Health and Social Affairs, Government of the Federated States of Micronesia, Palikir, Federated States of Micronesia. [194]Occupational and Environmental Epidemiology Section, Cancer Prevention and Research Institute, Florence, Italy. [195]Postgraduate Program in Epidemiology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil. [196]Department of Primary Care and Public Health, Imperial College London, London, UK. [197]Health Improvement Directorate, Public Health England, London, UK. [198]School of Public Health, University of Haifa, Haifa, Israel. [199]School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. [200]Department of Epidemiology and Biostatistics, Zhengzhou University, Zhengzhou, China. [201]Department of Pharmacology, Tehran University of Medical Sciences, Tehran, Iran. [202]Obesity Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [203]Department of Radiology, Johns Hopkins University, Baltimore, MD, USA. [204]Global and Community Mental Health Research Group, University of Macau, Macao, China. [205]School of Health and Environmental Studies, Hamdan Bin Mohammed Smart University, Dubai, United Arab Emirates. [206]Tabriz University of Medical Sciences, Tabriz, Iran. [207]Biomedical Research Networking Center for Mental Health Network (CIBERSAM), Madrid, Spain. [208]Research and Development Unit, San Juan de Dios Sanitary Park, Sant Boi De Llobregat, Spain. [209]School of Nutrition and Food Sciences, Tabriz University of Medical Sciences, Tabriz, Iran. [210]Healthcare Management, Maragheh University of Medical Sciences, Maragheh, Iran. [211]Department of Microbiology, Tehran University of Medical Sciences, Tehran, Iran. [212]Department of Microbiology, Maragheh University of Medical Sciences, Maragheh, Iran. [213]School of Nursing and Midwifery Tabriz University of Medical Sciences, Tabriz, Iran. [214]Independent Consultant, Tabriz, Iran. [215]Public Health Department, Mizan-Tepi University, Teppi, Ethiopia. [216]Unit of Epidemiology and Social Medicine, University Hospital Antwerp, Antwerp, Belgium. [217]School of Public Health, Curtin University, Bentley, Western Australia, Australia. [218]Medical Biology Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran. [219]Population Health, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia. [220]Center of Excellence in Behavioral Medicine, Nguyen Tat Thanh University, Ho Chi Minh, Vietnam. [221]Social Determinants of Health Research Center, Guilan Road Trauma Research Center, Guilan University of Medical Sciences, Rasht, Iran. [222]Guilan Road Trauma Research Center, Guilan University of Medical Sciences, Rasht, Iran. [223]Radiology and Nuclear Medicine Department, Kermanshah University of Medical Sciences, Kermanshah, Iran. [224]Department of Pharmacology and Therapeutics, University of Dhaka, Dhaka, Bangladesh. [225]Department of Epidemiology and Biostatistics, Tehran University of Medical Sciences, Tehran, Iran. [226]Computer Science Department, University of Human Development, Sulaimaniyah, Iraq. [227]Department of Internal Medicine, Bucharest Emergency Hospital, Bucharest, Romania. [228]Faculty of Dentistry, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [229]Clinical Legal Medicine, National Institute of Legal Medicine Mina Minovici, Bucharest, Romania. [230]Division of Information and Computing Technology, Hamad Bin Khalifa University, Doha, Qatar. [231]Qatar Foundation for Education, Science and Community Development, Doha, Qatar. [232]Faculty of Medicine Tunis, Medicine School of Tunis, Baab Saadoun, Tunisia. [233]Department of Community Medicine, University of Ibadan, Ibadan, Nigeria. [234]Department of Public Health, Lorestan University of Medical Sciences, Khorramabad, Iran. [235]Global Health and Development Department, Taipei Medical University, Taipei City, Taiwan. [236]Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [237]MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. [238]Harvard University, Boston, MA, USA. [239]Institute for Physical Activity and Nutrition, Deakin University, Burwood, Victoria, Australia. [240]Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia. [241]Institute of Family Medicine and Public Health, University of Tartu, Tartu, Estonia. [242]Psychosis Department, Babol Nushirvani University of Technology, Babol, Iran. [243]Psychiatric Department, Kermanshah University of Medical Sciences, Kermanshah, Iran. [244]Department of Medical Mycology, Mazandaran University of Medical Sciences, Sari, Iran. [245]Faculty of Graduate Studies, University of Colombo, Colombo, Sri Lanka. [246]Institute of Medicine, University of Colombo, Colombo, Sri Lanka. [247]School of Midwifery, A.T. Still University, Mesa, AZ, USA. [248]Environmental Research Center, Duke Kunshan University, Kunshan, China. [249]Department of Medicine, University of Miami, Atlantis, FL, USA. [250]Department of Ophthalmology, Heidelberg University, Heidelberg, Germany. [251]Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Beijing, China. [252]Social Determinants of Health Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. [253]Faculty of Medicine and Health Sciences, University of Opole, Opole, Poland. [254]Department of Family Medicine and Public Health, University of Opole, Opole, Poland. [255]Minimally Invasive Surgery Research Center, Iran University of Medical Sciences, Tehran, Iran. [256]Department of Nutrition and Dietetics, Mekelle University, Mekelle, Ethiopia. [257]Mazandaran University of Medical Sciences, Sari, Iran. [258]Isfahan University of Medical Sciences, Isfahan, Iran. [259]Social Determinants of Health Research Center, Qazvin University of Medical Sciences, Qazvin, Iran. [260]Department of Epidemiology, Hamadan University of Medical Sciences, Hamadan, Iran. [261]Research and Development, Australian Red Cross Blood Service, Sydney, New South Wales, Australia. [262]School of Public Health and Community Medicine, University of New South Wales, Sydney, New South Wales, Australia. [263]Hematologic Malignancies Research Center, Tehran University of Medical Sciences, Tehran, Iran. [264]Hematology-Oncology and Stem Cell Transplantation Research Center, Tehran University of Medical Sciences, Tehran, Iran. [265]Department of Anesthesiology & Pain Medicine, University of Washington, Seattle, WA, USA. [266]Odel Campus, University of Nairobi, Nairobi, Kenya. [267]Michigan State University, East Lansing, MI, USA. [268]Tabriz Health Management Research Center, Tabriz University of Medical Sciences, Tabriz, Iran. [269]National Institute for Health Research (NIHR), Tehran University of Medical Sciences, Tehran, Iran. [270]Department of Public Health and Community Medicine, Jordan University of Science and Technology, Ramtha, Jordan. [271]Social Determinants of Health Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. [272]Epidemiology and Biostatistics Department, Health Services Academy, Islamabad, Pakistan. [273]Population Studies, International Institute for Population Sciences, Mumbai, India. [274]Department of Internal Medicine, John H. Stroger Jr Hospital of Cook County, Chicago, IL, USA. [275]Department of Internal Medicine, Dow University of Health Sciences, Karachi, Pakistan. [276]Institute of Health Policy and Management, Seoul National University, Seoul, South Korea. [277]Department of Health Policy and Management, Seoul National University, Seoul, South Korea. [278]Faculty of Health and Wellbeing, Sheffield Hallam University, Sheffield, UK. [279]Department of Arts and Sciences, Ohio University, Zanesville, OH, USA. [280]Internal Medicine and Gastroenterology Department, National Hepatology and Tropical Research Institute, Cairo, Egypt. [281]Department of Medical Parasitology, Cairo University, Cairo, Egypt. [282]Department of Environmental Health Engineering, Hamadan University of Medical Sciences, Hamadan, Iran. [283]Department of Public Health, Mazandaran University of Medical Sciences, Sari, Iran. [284]Department of Nutrition and Health Science, Ball State University, Muncie, IN, USA. [285]School of Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. [286]School of Medicine, Xiamen University Malaysia, Sepang, Malaysia. [287]Department of Nutrition, Simmons College, Boston, MA, USA. [288]Department of Health Management and Health Economics, Kristiania University College, Oslo, Norway. [289]Department of Health Services Policy and Management, University of South Carolina, Columbia, SC, USA. [290]Nursing and Health Promotion, Oslo Metropolitan University, Oslo, Norway. [291]Department of Public Health, Debre Berhan University, Debre Berhan, Ethiopia. [292]Independent Consultant, Jakarta, Indonesia. [293]Department of Internal and Pulmonary Medicine, Sheri Kashmir Institute of Medical Sciences, Srinagar, India. [294]CIBERSAM, San Juan de Dios Sanitary Park, Sant Boi De Llobregat, Spain. [295]Department of Zoology, University of Oxford, Oxford, UK. [296]Medical School, Harvard University, Boston, MA, USA. [297]Department of Anthropology, Panjab University, Chandigarh, India. [298]Family and Community Health, University of Health and Allied Sciences, Ho, Ghana. [299]Psychology and Health Promotion, University of Kwazulu-Natal, Durban, South Africa. [300]Department of Psychiatry, University of Nairobi, Nairobi, Kenya. [301]Department of Psychology, University College London, London, UK. [302]International Institute for Population Sciences, Mumbai, India. [303]Department of Public Health Medicine, University of Kwazulu-Natal, Durban, South Africa. [304]Nursing, St John of God Hospital, Duayaw Nkwanta, Ghana. [305]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [306]Oxford Biomedical Research Centre, National Institute for Health Research (NIHR), Oxford, UK. [307]Department of Pediatrics, Post Graduate Institute of Medical Education and Research, Chandigarh, India. [308]Department of Community and Family Medicine, Academy of Medical Science, Baghdad, Iraq. [309]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [310]Department of Clinical Chemistry and Pharmacology, Uppsala University Hospital, Uppsala, Sweden. [311]School of Nursing, Hong Kong Polytechnic University, Hong Kong, China. [312]Department of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [313]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, China. [314]Department of Dentistry, Radboud University, Nijmegen, The Netherlands. [315]Section for Translational Health Economics, Heidelberg University Hospital, Heidelberg, Germany. [316]Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK. [317]Alliance for Improving Health Outcomes, Quezon City, The Philippines. [318]Institute of Nutrition, Friedrich Schiller University Jena, Jena, Germany. [319]Competence Cluster for Nutrition and Cardiovascular Health (NUTRICARD), Jena, Germany. [320]Physiology Department, Suez Canal University, Ismailia, Egypt. [321]Proteomics and Metabolomics Unit, Suez Canal University, Ismailia, Egypt. [322]Department of Cardiology, Damietta University, Damietta, Egypt. [323]Ophthalmology Department, Aswan Faculty of Medicine, Aswan, Egypt. [324]Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran. [325]Non-communicable Diseases Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. [326]Department of Maternal and Child Nursing and Public Health, Federal University of Minas Gerais, Belo Horizonte, Brazil. [327]Institute for Social Science Research, The University of Queensland, Brisbane, Queensland, Australia. [328]Ophthalmology Department, Iran University of Medical Sciences, Tehran, Iran. [329]Department Ophthalmology, University of Manitoba, Winnipeg, Manitoba, Canada. [330]Surgery Department, Emergency University Hospital Bucharest, Bucharest, Romania. [331]Department of Health Education and Health Promotion, Iran University of Medical Sciences, Tehran, Iran. [332]Campus Caucaia, Federal Institute of Education, Science and Technology of Ceará, Caucaia, Brazil. [333]Faculty of Health and Education, Botho University-Botswana, Gaborone, Botswana. [334]Division of Plastic Surgery, University of Washington, Seattle, WA, USA. [335]School of Medicine, University of New South Wales, Sydney, New South Wales, Australia. [336]Research Department, The George Institute for Global Health, New Delhi, India. [337]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. [338]Department of Health Services Research and Policy, London School of Hygiene & Tropical Medicine, London, UK. [339]Preventive Oncology Department, National Institute of Cancer Prevention and Research, Noida, India. [340]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA. [341]Peru Country Office, United Nations Population Fund (UNFPA), Lima, Peru. [342]Forensic Medicine Division, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. [343]Neurocenter, Helsinki University Hospital, Helsinki, Finland. [344]School of Health Sciences, University of Melbourne, Parkville, Victoria, Australia. [345]Breast Surgery Unit, Helsinki University Hospital, Helsinki, Finland. [346]University of Helsinki, Helsinki, Finland. [347]Clinical Microbiology and Parasitology Unit, Dr Zora Profozic Polyclinic, Zagreb, Croatia. [348]University Centre Varazdin, University North, Varazdin, Croatia. [349]Pacific Institute for Research & Evaluation, Calverton, MD, USA. [350]Health, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada. [351]Department of Computer Science and Software Engineering, University of Western Australia, Perth, Western Australia, Australia. [352]Department of Public Health, Amrita Institute of Medical Sciences, Kochi, India. [353]Department of Clinical Biochemistry, Babol University of Medical Sciences, Babol, Iran. [354]Golestan University of Medical Sciences, Gorgan, Iran. [355]Foodborne and Waterborne Diseases Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [356]Faculty of General Medicine, Kyrgyz State Medical Academy, Bishkek, Kyrgyzstan. [357]Department of Atherosclerosis and Coronary Heart Disease, National Center of Cardiology and Internal Disease, Bishkek, Kyrgyzstan. [358]Health Equity Research Center, Tehran University of Medical Sciences, Tehran, Iran. [359]Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden. [360]Department of Food Technology, College of Agriculture, Salahaddin University-Erbil, Erbil, Iraq. [361]Information Technology Department, University of Human Development, Sulaimaniyah, Iraq. [362]Department of Biostatistics, Hamadan University of Medical Sciences, Hamadan, Iran. [363]School of Pharmacy, Haramaya University, Harar, Ethiopia. [364]Institute of Public Health, Heidelberg University, Heidelberg, Germany. [365]Health Systems and Policy Research Unit, Ahmadu Bello University, Zaria, Nigeria. [366]Faculty of Life Sciences and Medicine, King's College London, London, UK. [367]Clinical Epidemiology and Public Health Research Unit, Burlo Garofolo Institute for Maternal and Child Health, Trieste, Italy. [368]Health Sciences Research Center, Mazandaran University of Medical Sciences, Sari, Iran. [369]Department of Epidemiology and Biostatistics, Kurdistan University of Medical Sciences, Sanandaj, Iran. [370]Social Determinants of Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran. [371]Department of Epidemiology, Iran University of Medical Sciences, Tehran, Iran. [372]Department of Mathematical Sciences, University of Bath,

Bath, UK. [373]International Laboratory for Air Quality and Health, Queensland University of Technology, Brisbane, Queensland, Australia. [374]Department of Clinical Biochemistry, Tarbiat Modares University, Tehran, Iran. [375]Department of Health Management and Economics, Tehran University of Medical Sciences, Tehran, Iran. [376]Federal Institute for Population Research, Wiesbaden, Germany. [377]Center for Population and Health, Wiesbaden, Germany. [378]Department of Epidemiology and Biostatistics, University of Gondar, Gondar, Ethiopia. [379]Department of Pediatric Medicine, Nishtar Medical University, Multan, Pakistan. [380]Department of Pediatrics & Pediatric Pulmonology, Institute of Mother & Child Care, Multan, Pakistan. [381]Department of Urology, Tehran University of Medical Sciences, Tehran, Iran. [382]Operating Room Department, Kermanshah University of Medical Sciences, Kermanshah, Iran. [383]Research and Analytics, Initiative for Financing Health and Human Development, Chennai, India. [384]Research and Analytics, Bioinsilico Technologies, Chennai, India. [385]Cancer Research Center of Cancer Institute, Tehran University of Medical Sciences, Tehran, Iran. [386]Department of Epidemiology & Biostatistics, Kermanshah University of Medical Sciences, Kermanshah, Iran. [387]Suraj Eye Institute, Nagpur, India. [388]Cochrane South Africa, South African Medical Research Council, Cape Town, South Africa. [389]Emergency Hospital of Bucharest, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [390]General Surgery Department, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [391]Anatomy and Embryology Department, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. [392]Department of Cardiology, Cardio-aid, Bucharest, Romania. [393]Department of Biological Sciences, University of Embu, Embu, Kenya. [394]Institute for Global Health Innovations, Duy Tan University, Hanoi, Vietnam. [395]Center for Excellence in Behavioral Health, Nguyen Tat Thanh University, Ho Chi Minh, Vietnam. [396]Global Health Department, University of Washington, Seattle, WA, USA. [397]Department of Pediatrics, University of Washington, Seattle, WA, USA. [398]Clinical Pharmacy Unit, Mekelle University, Mekelle, Ethiopia. [399]Public Health Science Department, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa. [400]Department of Community and Family Medicine, Iran University of Medical Sciences, Tehran, Iran. [401]University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. [402]Department of Health Economics, Tabriz University of Medical Sciences, Tabriz, Iran. [403]Department of Medicine, University of Cape Town, Cape Town, South Africa. [404]Centre of Cardiovascular Research and Education in Therapeutics, Monash University, Melbourne, Victoria, Australia. [405]Independent Consultant, Accra, Ghana. [406]Translational Health Research Institute, Western Sydney University, Penrith, New South Wales, Australia. [407]Center for the Aid Program of Research in South Africa (CAPRISA) TB and HIV Pathogenesis Unit, United Nations Programme on HIV/AIDS (UNAIDS), Durban, South Africa. [408]Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada. [409]Department of Psychiatry, University of Lagos, Lagos, Nigeria. [410]Gastroenterology and Liver Disease Research Center, A.C.S. Medical College and Hospital, Tehran, Iran. [411]Department of Food Science and Postharvest Technology, Gulu University, Gulu, Uganda. [412]Ghent University, Ghent, Belgium. [413]Centre for Healthy Start Initiative, Lagos, Nigeria. [414]Department of Health Promotion and Education, University of Ibadan, Ibadan, Nigeria. [415]Department of Pharmacology and Therapeutics, University of Nigeria Nsukka, Enugu, Nigeria. [416]University of Washington, Seattle, WA, USA. [417]Graduate School of Public Health, San Diego State University, San Diego, CA, USA. [418]Center for Health Systems Research, National Institute of Public Health, Cuernavaca, Mexico. [419]School of Medicine, Autonomous University of Madrid, Madrid, Spain. [420]Department of Nephrology and Hypertension, The Institute for Health Research Foundation Jiménez Díaz University Hospital, Madrid, Spain. [421]Environmental Management and Toxicology, University of Benin, Benin City, Nigeria. [422]Faculty of Geoinformation Science and Earth Observation, University of Twente, Enschede, The Netherlands. [423]Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana. [424]Analytical Center, Moscow Institute of Physics and Technology, Dolgoprudny, Russia. [425]Committee for the Comprehensive Assessment of Medical Devices and Information Technology, Health Technology Assessment Association, Moscow, Russia. [426]Institute for Advanced Medical Research and Training, University of Ibadan, Ibadan, Nigeria. [427]Department of Tb & Respiratory Medicine, Jagadguru Sri Shivarathreeswara University, Mysore, India. [428]Department of Medicine, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada. [429]Heidelberg University, Heidelberg, Germany. [430]Department of Medical Humanities and Social Medicine, Kosin University, Busan, South Korea. [431]Research and Evaluation, Population Council, New Delhi, India. [432]Indian Institute of Health Management Research University, Jaipur, India. [433]Center for Research and Innovation, Ateneo De Manila University, Pasig City, The Philippines. [434]Department of Genetics, Harvard University, Boston, MA, USA. [435]Laboratory of Genetics and Molecular Cardiology, University of São Paulo, Sao Paulo, Brazil. [436]Department of Cardiology, University of Bern, Bern, Switzerland. [437]Parasitology and Entomology Department, Tarbiat Modares University, Tehran, Iran. [438]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [439]Department of Nutrition and Food Sciences, Maragheh University of Medical Sciences, Maragheh, Iran. [440]Department of Public Health, Maragheh University of Medical Sciences, Maragheh, Iran. [441]Non-communicable Diseases Research Center, Alborz University of Medical Sciences, Karaj, Iran. [442]Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran. [443]College of Graduate Health Sciences, A.T. Still University, Mesa, AZ, USA. [444]Medichem, Barcelona, Spain. [445]Molecular and Cell Biology Research Center, Mazandaran University of Medical Sciences, Sari, Iran. [446]Department of Immunology, Mazandaran University of Medical Sciences, Sari, Iran. [447]Endocrinology and Metabolism Molecular-Cellular Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran. [448]Thalassemia and Hemoglobinopathy Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. [449]Department of Clinical Biochemistry, Kermanshah University of Medical Sciences, Kermanshah, Iran. [450]Sina Trauma and Surgery Research Center, Tehran University of Medical Sciences, Tehran, Iran. [451]Department of Health Education & Promotion, Kermanshah University of Medical Sciences, Kermanshah, Iran. [452]Department of Nephrology, Nizam's Institute of Medical Sciences, Hyderabad, India. [453]Prevention of Metabolic Disorders Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [454]Critical Care Quality Improvement Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [455]Policy Research Institute, Kathmandu, Nepal. [456]Institute for Poverty Alleviation and International Development, Yonsei University, Wonju, South Korea. [457]Institute of Public Health, Federal University of Bahia, Salvador, Brazil. [458]Gonçalo Moniz Institute, Oswaldo Cruz Foundation, Salvador, Brazil. [459]School of Behavioral Sciences and Mental Health, Iran University of Medical Sciences, Tehran, Iran. [460]Social Science and Psychology, Western Sydney University, Penrith, New South Wales, Australia. [461]School of Social Sciences and Psychology, Western Sydney University, Penrith, New South Wales, Australia. [462]Research Center for Environmental Determinants of Health (RCEDH), Kermanshah University of Medical Sciences, Kermanshah, Iran. [463]Department of Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [464]Department of Epidemiology, Birjand University of Medical Sciences, Birjand, Iran. [465]EPIUnit, University of Porto, Porto, Portugal. [466]Department of Clinical Research, Federal University of Uberlândia, Uberlândia, Brazil. [467]Public Health, Addis Ababa University, Addis Ababa, Ethiopia. [468]Department of Public Health, Wollega University, Nekemte, Ethiopia. [469]Martin School, University of Oxford, Oxford, UK. [470]Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Gorgan, Iran. [471]Epidemiology and Biostatistics, Kurdistan University of Medical Sciences, Sanandaj, Iran. [472]Infectious Diseases and Tropical Medicine Research Center, Babol University of Medical Sciences, Babol, Iran. [473]School of Biotechnology, Ikiam Amazon Regional University, Tena, Ecuador. [474]Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China. [475]Department of Biomedical Sciences, University of Sassari, Sassari, Italy. [476]Department of Health, Safety and Environment (HSE), Shahid Beheshti University of Medical Sciences, Tehran, Iran. [477]Faculty of Public Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. [478]Department of Neuroscience, Iran University of Medical Sciences, Tehran, Iran. [479]Neurogenic Inflammation Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. [480]Biotechnology Research Center, Mashhad University of Medical Sciences, Mashhad, Iran. [481]Department of Anatomical Sciences, Kermanshah University of Medical Sciences, Kermanshah, Iran. [482]Department of Pathology, Al-Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. [483]School of Health and Policy Management, Faculty of Health, York University, Toronto, Ontario, Canada. [484]Taleghani Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran. [485]Department of Radiology and Nuclear Medicine, Kermanshah University of Medical Sciences, Kermanshah, Iran. [486]Taleghani Hospital, Kermanshah, Iran. [487]Center for Health Policy & Center for Primary Care and Outcomes Research, Stanford University, Stanford, CA, USA. [488]Department of Entomology, Ain Shams University, Cairo, Egypt. [489]Centre School of Public Health and Health Management, University of Belgrade, Belgrade, Serbia. [490]Post-graduate Program in Infectious Diseases and Tropical Medicine, Federal University of Minas Gerais, Belo Horizonte, Brazil. [491]Department of Community Medicine, PSG Institute of Medical Sciences and Research, Coimbatore, India. [492]PSG-FAIMER South Asia Regional Institute, Coimbatore, India. [493]Department of Health and Society, Faculty of Medicine, University of Applied and Environmental Sciences, Bogotá, Colombia. [494]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. [495]Surgery Department, Hamad Medical Corporation, Doha, Qatar. [496]Faculty of Health & Social Sciences, Bournemouth University, Bournemouth, UK. [497]School of Public Health, Imperial College London, London, UK. [498]Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, USA. [499]Center of Expertise in Microbiology, Tehran University of Medical Sciences, Tehran, Iran. [500]Invasive Fungi Research Center, Mazandaran University of Medical Sciences, Sari, Iran. [501]Department of Health Promotion and Education, Alborz University of Medical Sciences, Karaj, Iran. [502]Independent Consultant, Karachi, Pakistan. [503]School of Medicine, Dezful University of Medical Sciences, Dezful, Iran. [504]Medical Laboratory Sciences, Mazandaran University of Medical Sciences, Sari, Iran. [505]Chronic Diseases (Home Care) Research Center, Hamadan University of Medical Sciences, Hamadan, Iran. [506]Department of Laboratory Sciences, Karaj Islamic Azad University, Kermanshah, Iran. [507]Department of Basic Sciences, Karaj Islamic Azad University, Kermanshah, Iran. [508]HIV/STI Surveillance Research Center, Kerman University of Medical Sciences, Kerman, Iran. [509]Policy and Planning Division, Ministry of Health, Riyadh, Saudi Arabia. [510]University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India. [511]Division of General Internal Medicine and Primary Care, Harvard University, Boston, MA, USA. [512]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [513]National Institute of Infectious Diseases, Tokyo, Japan. [514]Finnish Institute of Occupational Health, Helsinki, Finland. [515]Institute of Medical Epidemiology, Martin Luther University Halle-Wittenberg, Halle, Germany. [516]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. [517]Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [518]Department of Epidemiology, School of Preventive Oncology, Patna, India. [519]Department of Epidemiology, Healis Sekhsaria Institute for Public Health, Mumbai, India. [520]Department of Physiotherapy and Occupational Therapy, Næstved-Slagelse-Ringsted Hospitals, Slagelse, Denmark. [521]Medical Division, German Leprosy and TB Relief Association Ethiopia, Addis Ababa, Ethiopia. [522]Department of Medicine, University of Washington, Seattle, WA, USA. [523]Department of Medicine, University of Calgary, Calgary, Alberta, Canada. [524]Social Development and Health Promotion Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran. [525]Hospital Universitario de la Princesa, Autonomous University of Madrid, Madrid, Spain. [526]Centro de Investigación en Red de Enfermedades Respiratorias (CIBERES), Institute of Health Carlos III, Madrid, Spain. [527]Division of Community Medicine, International Medical University, Kuala Lumpur, Malaysia. [528]Department of Nursing, Muhammadiyah University of Surakarta, Kartasura, Indonesia. [529]Department of Community Medicine, Ahmadu Bello University, Zaria, Nigeria. [530]Department of Criminology, Law and Society, University of California Irvine, Irvine, CA, USA. [531]Neurology Department, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India. [532]Carlos III Health Institute, Biomedical Research Networking Center for Mental Health Network (CIBERSAM), Madrid, Spain. [533]Department of Medicine, University of Valencia, Valencia, Spain. [534]School of Social Work, University of Illinois, Urbana, IL, USA. [535]Cancer Control Center, Osaka International Cancer Institute, Osaka, Japan. [536]University Institute 'Egas Moniz', Monte Da Caparica, Portugal. [537]Research Institute for Medicines, Faculty of Pharmacy of Lisbon, University of Lisbon, Lisbon, Portugal. [538]Department of Pediatrics, King Saud University, Riyadh, Saudi Arabia. [539]College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. [540]Anesthesiology Department, University of

Virginia, Charlottesville, VA, USA. [541]Syrian Expatriate Medical Association (SEMA), Charlottesville, VA, USA. [542]Department of Public Health and Community Medicine, Central University Kerala, Kasaragod, India. [543]Nanyang Technological University, Singapore, Singapore. [544]School of Exercise and Nutrition Sciences, Queensland University of Technology, Brisbane, Queensland, Australia. [545]Department of Pathology and Legal Medicine, University of São Paulo, Sao Paulo, Brazil. [546]Department of Health Economics, Hanoi Medical University, Hanoi, Vietnam. [547]Department of Molecular Medicine and Pathology, University of Auckland, Auckland, New Zealand. [548]Clinical Hematology and Toxicology, Military Medical University, Hanoi, Vietnam. [549]Gomal Center of Biochemistry and Biotechnology, Gomal University, Dera Ismail Khan, Pakistan. [550]TB Culture Laboratory, Mufti Mehmood Memorial Teaching Hospital, Dera Ismail Khan, Pakistan. [551]Division of Health Sciences, University of Warwick, Coventry, UK. [552]Department of Education and Health, Trauma Research Center, Tehran, Iran. [553]Critical and Intensive Care Department, Trauma Research Center, Tehran, Iran. [554]Argentine Society of Medicine, Buenos Aires, Argentina. [555]Velez Sarsfield Hospital, Buenos Aires, Argentina. [556]University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [557]Department of General Practice, University Medical Center Groningen, Groningen, The Netherlands. [558]Ukk Institute, Tampere, Finland. [559]Psychosocial Injuries Research Center, Ilam University of Medical Sciences, Ilam, Iran. [560]Raffles Neuroscience Centre, Raffles Hospital, Singapore, Singapore. [561]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [562]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy. [563]Occupational Health Unit, Sant'orsola Malpighi Hospital, Bologna, Italy. [564]Department of Information Technologies and Management, Moscow Institute of Physics and Technology, Dolgoprudny, Russia. [565]Department of Information and Internet Technologies, I. M. Sechenov First Moscow State Medical University, Moscow, Russia. [566]Department of Health Care Administration and Economy, National Research University Higher School of Economics, Moscow, Russia. [567]Foundation University Medical College, Foundation University, Rawalpindi, Pakistan. [568]Department of Statistics, University of Washington, Seattle, WA, USA. [569]Department of Epidemiology and Biostatistics, Wuhan University, Wuhan, China. [570]Department of Psychiatry, University of São Paulo, Sao Paulo, Brazil. [571]Institute of Child Health, University College London, London, UK. [572]Cardiology Department, Royal Children's Hospital, Melbourne, Victoria, Australia. [573]Murdoch Childrens Research Institute, Melbourne, Victoria, Australia. [574]School of Nursing, Aksum University, Aksum, Ethiopia. [575]Competence Center of Mortality-Follow-Up, Federal Institute for Population Research, Wiesbaden, Germany. [576]Cochrane South Africa, Medical Research Council South Africa, Cape Town, South Africa. [577]Department of Global Health, Stellenbosch University, Cape Town, South Africa. [578]Department of Pharmacology and Toxicology, Mekelle University, Mekelle, Ethiopia. [579]Department of Pharmacology, Addis Ababa University, Addis Ababa, Ethiopia. [580]Zhejiang Spine Research Center, Wenzhou Medical University, Wenzhou, China. [581]School of Medicine, Nanjing University, Nanjing, China. [582]Department of Diabetes and Metabolic Diseases, University of Tokyo, Tokyo, Japan. [583]Department of Health Management, Policy and Economics, Kerman University of Medical Sciences, Kerman, Iran. [584]Health Services Management Research Center, Kerman University of Medical Sciences, Kerman, Iran. [585]Department of Pediatrics, University of Jos, Jos, Nigeria. [586]Department of Pediatrics, Jos University Teaching Hospital, Jos, Nigeria. [587]Centre for Suicide Research and Prevention, University of Hong Kong, Hong Kong, China. [588]Department of Social Work and Social Administration, University of Hong Kong, Hong Kong, China. [589]Department of Psychopharmacology, National Center of Neurology and Psychiatry, Tokyo, Japan. [590]Health Economics & Finance, Global Health, Jackson State University, Jackson, MS, USA. [591]Research Center for Public Health, Tsinghua University, Peking, China. [592]Prevention of Cardiovascular Disease Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [593]Medical Parasitology Department, Cairo University, Cairo, Egypt. [594]Global Health Institute, Wuhan University, Wuhan, China. [595]Department of Health Management and Economics, A.C.S. Medical College and Hospital, Tehran, Iran. [596]Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran. [597]Electrical Engineering, Institute for Research in Fundamental Sciences, Tehran, Iran. [598]Social Determinants of Health Research Center, Ardabil University of Medical Science, Ardabil, Iran. [599]Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research Bangladesh, Dhaka, Bangladesh. [600]Department of Medicine, Monash University, Melbourne, Victoria, Australia. [601]Student Research Committee, Babol University of Medical Sciences, Babol, Iran. [602]Department of Community Medicine, Ardabil University of Medical Science, Ardabil, Iran. [603]Social Development and Health Promotion Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran. [604]Maternal and Child Wellbeing Unit, African Population Health Research Centre, Nairobi, Kenya. [605]Public Health Department, Dilla University, Dilla, Ethiopia. [606]Department of Preventative Medicine, Wuhan University, Wuhan, China. [607]School of Public Health, Wuhan University of Science and Technology, Wuhan, China. [608]Indian Institute of Public Health, Public Health Foundation of India, Gurugram, India. [609]School of BioSciences, University of Melbourne, Parkville, Victoria, Australia. [610]State University of Semarang, Public Health Science Department, Kota Semarang, Indonesia. [611]Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei City, Taiwan. [612]These authors contributed equally: Roy Burstein, Nathaniel J. Henry. [613]These authors jointly supervised this work: Christopher J. L. Murray, Simon I. Hay. *e-mail: sihay@uw.edu

# METHODS

**Overview.** We fitted a discrete hazards geostatistical model[51,52] with correlated space–time–age errors and made predictions to generate joint estimates—with uncertainty—of the probability of death (the number of deaths per live births) and the number of deaths for children aged 0–28 days (neonates), children under 1 year old (infants) and children under 5 years old at the subnational level for 99 LMICs for each year from 2000 to 2017. The analytical process is summarized in the flowchart in Extended Data Fig. 6. We made estimates at a grid-cell resolution of approximately 5 × 5-km and then produced spatially aggregated estimates at the first (that is, states or provinces) and second (that is, districts or counties) administrative levels, as well as the country level.

Countries were selected for inclusion in this study based on their socio-demographic index (SDI) published in the Global Burden of Disease study (GBD)[53]. The SDI is a measure of development based on income per capita, educational attainment and fertility rates among women under 25 years old. We primarily aimed to include all countries in the middle, low–middle or low SDI quintiles, with several exceptions. Brazil and Mexico were excluded despite middle SDI status owing to the availability of high-quality vital registration data in these countries, which have served as the basis for existing subnational estimates of child mortality. Because this study did not incorporate vital registration data sources (see 'Limitations'), Brazil and Mexico were not estimated directly; instead, state-level estimates from the GBD 2017 study were directly substituted in figures where appropriate[4]. Albania and Moldova were excluded despite middle SDI status owing to geographical discontinuity with other included countries and lack of available survey data. North Korea was excluded despite low–middle SDI status owing to geographical discontinuity and insufficient data. As countries with high–middle SDI status in 2017, China and Malaysia were excluded from this analysis. Libya was included despite high–middle SDI status to create better geographical continuity. Island nations with populations under 1 million were excluded because they typically lacked sufficient survey data or geographical continuity for a geospatial analytic approach to be advantageous over a national approach. Supplementary Figure 3.1 shows a map of the countries included in this study and Supplementary Table 3.1 lists the countries.

**Data.** We extracted individual records from 555 household sample survey and census sources. Records came in the form of either summary birth histories (SBHs) or complete birth histories (CBHs). All input data were subject to quality checks, which resulted in the exclusion of 82 surveys and censuses owing to quality concerns (see Supplementary Information section 3.2 for more details). Data on life and mortality experiences from CBH sources can be tabulated directly into discrete period and age bins, thus allowing for period-specific mortality estimations, known as the synthetic cohort method[54–56]. For SBH data, we used indirect estimation[57] to estimate age-specific mortality probabilities and sample sizes and assign them to specific time periods. Complete details are available in Supplementary Information section 3.3.2.

In all cases, after pre-processing, each data point provided a number of deaths and a sample size for an age bin in a specific year and location. We referenced all data points to GPS coordinates (latitude and longitude) wherever possible. In cases in which GPS data were unavailable, we matched data points to the smallest possible areal unit (also referred to as 'polygons'). All polygon data were spatially resampled into multiple GPS coordinates and weighted based on the population distribution following a previously described procedure[5,22,23,58] and described in Supplementary Information section 3.4. Our combined global dataset contained approximately 15.9 million births and 1.1 million child deaths. A complete list of data sources is provided in Supplementary Table 8.1.

In addition to data on child mortality, we used a number of spatial data sources for this analysis. These included a suite of geospatial covariates, population estimates and administrative boundaries[68]. These sources and processing procedures are described in Supplementary Information section 4.

**Spatial covariates.** We extracted values from each of 10 geospatial covariates at each data point location. Geospatial covariates are spatial data represented at the 5 × 5-km grid-cell resolution. The covariates were travel time to the nearest city, educational attainment of maternal-aged women, the ratio of population of children under 5 to women of reproductive age (ages 15–49 years old), the mass per cubic meter of air of particles with a diameter less than 2.5 µm, total population, a binary indicator of urbanicity, intensity of lights at night, the proportion of children aged 12–23 months who had received the third dose of diphtheria–pertussis–tetanus vaccine, incidence rate of *Plasmodium falciparum*-associated malaria in children under 5 and prevalence of stunting in children under 5 (see Supplementary Information). All covariate values were centred on their means and scaled by their standard deviations. Covariates typically had global spatial coverage and values that vary by year. More details of the spatial covariates can be found in Supplementary Information section 4.

**Analysis.** *Geostatistical model.* To synthesize information across various sources, and to make consistent estimates across space and time, we fitted discrete hazards[51,52] geostatistical models[59] to our data. The models were discrete in the sense that ages were represented in seven mutually exclusive bins (0, 1–5, 6–11, 12–23, 24–35, 36–47 and 48–59 months), each with its own assumed constant mortality probability. The model explicitly accounted for variation across age bin, year and space through inclusion of both fixed and random effects. Indicator variables for each age bin were included to form a discrete baseline mortality hazard function, representing the risk of mortality in discrete bins from birth to 59 months of age with covariates set at their means. Baseline hazard functions were allowed to vary in space and time in response to changing covariate values, as well as in response to linear effect on year. To model this relationship, we estimated the effect of each covariate value on the risk of mortality. These estimated effects were then applied to the gridded surface of covariate values to make predictions across the entire study geography. We also included a Gaussian random effect across countries to account for larger-scale variations due to political or institutional effects, as well as a Gaussian random effect for each data source to account for source-specific biases. Finally, we included a Gaussian process random effect with a covariance matrix structured to account for remaining correlation across age, time and physical space. As such, estimates at a specific age, time or place benefitted from drawing predictive strength from data points nearby in all of these dimensions.

For each modelling region, we fitted one such discrete hazards model with a binomial data likelihood. All data were prepared such that we counted or estimated the number of children entering into ($n$) and dying within ($Y$) each period–age bin from each GPS-point location ($s$) in each survey ($k$) within each country ($c$).

The number of deaths for children in age band ($a$) in year ($t$) at location ($s$) was assumed to follow a binomial distribution:

$$Y_{a,s,t} \sim \text{binomial}(n_{a,s,t},\ P_{a,s,t})$$

where $P_{a,s,t}$ is the probability of death in age bin $a$, conditional on survival to that age bin for a particular space–time location. Using a generalized linear regression modelling framework, a logit link function is used to relate $P$ to a linear combination of effects:

$$\text{logit}(P_{a,s,t}) = \beta^0 + \sum_{a=2}^{7} I_a\beta_a^1 + \beta^2 X_{s,t} + \beta^3 t + \nu_{c\,[s]} + \nu_{k\,[s]} + Z_{a,s,t}$$

The first term, $\beta^0$, is an intercept, representing the mean for the first age band when all covariates are equal to zero, whereas $\beta_a^1$ are fixed effects for each age band, representing the mean overall hazard deviation for each age band from the intercept, when all other covariates are equal to zero. $\beta^2$ are the effects of geospatial covariates ($X_{s,t}$), which we describe in detail in Supplementary Information section 4. $\beta^3$ is an overall linear temporal effect to account for overall temporal trends within the region. All geospatial covariates were centred and scaled by subtracting their mean and dividing by their standard deviations. Each $\nu$ term represents uncorrelated Gaussian random effects: $\nu_{c\,[s]} \sim \text{normal}(0, \sigma_c^2)$ is a country-level random effect applied to all locations ($s$) within a country ($c$); $\nu_{k\,[s]} \sim \text{normal}(0, \sigma_k^2)$ is a data source-level random effect for the survey ($k$) from which the data at location $s$ were observed. Data source-level random effects were used to account for systematic variation or biases across data sources and were included in model fitting but not in prediction from fitted models. The term $Z_{a,s,t} \sim$ Gaussian process$(0, K)$ is a correlated random effect across age, space and time, and is modelled as a four-dimensional mean zero Gaussian process with covariance matrix $K$. This term accounts for structured residual correlation across these spatial–age–temporal dimensions that are not accounted for by any of the model's other fixed or random effects. This structure was chosen, because the hazard probability for each age group is expected to vary in space and time, and such spatiotemporal correlations are likely to be similar across ages. $K$ is constructed as a separable process across age, space and time $(K = \Sigma_a \otimes \Sigma_t \otimes \Sigma_s)$. The continuous spatial component is modelled with a stationary isotropic Matérn covariance function, and the age and temporal effects were each assumed to be discrete auto-regressive order 1. We provide further details on model fitting and specification in Supplementary Information section 5.1.

We assigned priors to all model parameters and performed maximum a posteriori inference using Template Model Builder[60] software in R version 3.4. We fitted the model separately for each of 11 world regions (see Supplementary Fig. 3.1), owing to memory constraints and to allow model parameters to vary across epidemiologically distinct regions.

*Post-estimation.* Using the joint precision matrix and point estimates, we generated 1,000 draws from all model parameters using a multivariate-normal approximation. These model parameter draws were used to predict corresponding draws of mortality probabilities across all age groups for each grid cell in each year. In other words, for each age bin in each year we estimated 1,000 gridded surfaces of mortality probability estimates, each surface corresponding to one draw from the posterior parameter estimates[61]. All subsequent post-estimation procedures were carried out across draws to propagate model uncertainty. We used these estimated spatiotemporal gridded surfaces of age-specific mortality probabilities to produce various final resulting data products.

From the fitted model parameters, we produced posterior mortality probability estimates for each age group for each 5 × 5-km grid cell for each year from 2000 to 2017. We combined gridded age group estimates to obtain infant (under 1) and child (under 5) mortality estimates at each gridded location. Using a conversion from mortality probability to mortality rates, and using a gridded surface of population, we also estimated the number of deaths that occurred in each age group at each location in each year. For both mortality probabilities and counts, we multiplied out corresponding gridded estimates by a constant to ensure that at the national—and in two countries, the first administrative-level unit—aggregated estimates for each age group and year were calibrated such that they equalled estimates in the GBD study[4]. This calibration allowed us to take advantage of national data sources, such as vital registration, that could not be used in this study. We also aggregated grid-cell-level estimates to first and second administrative-level units using gridded population surface to weight estimates. These steps are described in Supplementary Information section 5.2.

*Model validation.* We used fivefold cross-validation to assess and compare model performance with respect to estimating local trends of age-specific mortality. Each fold was created by combining complete surveys into subsets of approximately 20% of data sources from the input data. Holding out entire surveys at a time served as a comparable approximation to the type of missingness in our data, essentially helping us check how well our model estimates of mortality probabilities compared to empirical estimates of mortality probability from an unobserved data source that did not inform the model.

For each posterior draw, we aggregated to administrative units. Using data aggregated to the administrative unit and aggregated estimate pairs, we calculated the difference between out-of-sample empirical data estimates and modelled estimates, and we report the following summary metrics: mean error, which serves as a measure of bias, the square root of mean errors, which serves as a measure of the total variation in the errors, the correlation and 95% coverage. At the second administrative-level unit for under-5 mortality, our out-of-sample 95% coverage was 93%, correlation was 0.78, mean absolute error was 0.015 and mean error was −0.0011. These results indicate a good overall fit, with minimal bias. This procedure and the full validation results are discussed in Supplementary Information section 5.3.

**Limitations.** This work should be assessed in full acknowledgement of several data and methodological limitations. We exclusively used CBH and SBH data from household survey and census data sources. Ideally, estimates of child mortality should incorporate all available data, including data from administrative vital registration systems. Vital registration systems are commonly present in many middle-income and all high-income countries. There are known data-quality issues with vital registration sources in many middle-income countries[48,62] that add complications to their inclusion in our modelling procedure. For example, systems may not capture all deaths, and this level of 'underreporting' probably varies in space, time and age. In addition, underreporting is probably negatively correlated with mortality, and could contribute substantial bias to estimates. Statistical methods must be developed to jointly estimate—and adjust for—underreporting in vital registration data before such data can be used in geospatial models of child mortality. Promising work has begun in this domain in specific countries[63], but further advancement will be necessary to improve estimates across a time series and across many countries at once.

We assume that SBH and CBH data were retrospectively representative in the locations in which they were collected. As such, we assume that survey respondents did not migrate. High-spatial-resolution migration estimates with which to adjust estimates do not yet exist, and many of the data sources that we use do not collect information on migration. We conducted a focused sensitivity analysis (Supplementary Information section 5.4.4) for migration in six countries, and found that although our results were generally robust, there was variation by country. Furthermore, despite providing high-quality retrospective data from representative samples of households, birth history data can suffer from certain non-sampling issues, such as survival/selection biases[64] and misplacement of births[65]. We did not attempt to make corrections to data, and they were used as-is. Furthermore, retrospective birth history data will—by design—have a changing composition of maternal ages depending on the time since the survey. This was minimized by limiting retrospective trends to up to 17 years.

Although we collated a large geo-referenced database of survey data on child mortality, these data represented about 1% (1.1 million) of total deaths of children under 5 in study areas over the period. Where data do not exist or are not available in certain locations, mean estimates are informed from smoothing to nearby estimates and covariates. As such, there could be additional small-scale heterogeneity that is not picked up by our model. Wider uncertainty intervals in areas with no data account for these potential unknowns, and our 95% coverage estimates in out-of-sample predictive tests appear to be well-calibrated at the second administrative unit level. Furthermore, discrete localized mortality shock events could be missing in our analysis due to the lack of data and selection biases in surveys and censuses, and spatiotemporal smoothing. Fatal discontinuities are explicitly accounted for at the national or province level by calibration to GBD estimates. In all, 0.35% (0.4 million) of the 123 million deaths over this period were attributed to fatal discontinuities.

On the modelling side, we integrated point and areal data into a continuous model by constructing pseudo-points from areal data. Modelling approaches that integrate point and areal data as part of a joint model likelihood function are in development[66] but are currently computationally infeasible at the large geographical scales at which we currently model. Furthermore, we divided our models into 11 regional fits (see Supplementary Fig. 3.1), as a full model that encompasses all 99 countries would be computationally infeasible due to memory constraints. Splitting up modelling in this way had the benefit of enabling parameters to vary across epidemiologically distinct world regions. A preferred model, however, would be fitted to all data simultaneously with parameters that are spatially variable.

The separable model used for age–space–time correlations is a common parsimonious assumption afforded in applying spatiotemporal geostatistical models due to efficient computation and inference; however, it yields the assumption of fully symmetric covariance. The symmetry implicit in the separable model dictates, for example, that (holding age constant for simplicity) the covariance between the observations at (location 1, time 1) and (location 2, time 2) is the same as the covariance between (location 1, time 2) and (location 2, time 1). Given our available data density in space–age–time, we believe that attempting to parameterize a more complex non-separable model would be challenging both computationally and inferentially, and it is not clear whether there would be much to benefit from the extra complications.

There are several limitations to address with respect to the use of covariates in the model. Most of the geospatial covariates that we used in the geostatistical model were themselves estimates produced from various geospatial models. Some of those estimated surfaces used covariates that were also included in our model in their estimation process. As such, we emphasize that our model is meant to be predictive, and that drawing inference from fitted coefficients across these highly correlated covariates is problematic and not recommended. Furthermore, we assumed no measurement error in the covariate values and assumed that the functional form between mortality and all covariates was linear in logit space. In certain locations, we used covariate values for prediction that were outside the observed range of the training data. As we explore in Supplementary Information section 5.4.2, however, these areas represent a relatively small proportion of the population.

Finally, we used a method for indirect estimation of SBHs that was recently developed and validated[57]. As such, indirect estimation was carried out as a pre-processing step before fitting the geostatistical model. We attempted to propagate various forms of uncertainty that could be introduced in this step, which resulted in halving the total effective sample size across all SBH data. In future, we aim to fully integrate such processing into the statistical model; such methods are in development[67], but are not yet computationally feasible at scale.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The findings of this study are supported by data that are available from public online repositories, data that are publicly available upon request of the data provider and data that are not publicly available due to restrictions by the data provider. Non-publicly available data were used under a license for the current study, but may be available from the authors upon reasonable request and with permission of the data provider. A detailed table of data sources and availability can be found in Supplementary Table 8.1. The full output of the analyses is publicly available in the Global Health Data Exchange (GHDx; http://ghdx.healthdata.org/record/ihme-data/lmic-under5-mortality-rate-geospatial-estimates-2000-2017) and can be explored using custom data visualization tools (https://vizhub.healthdata.org/lbd/under5).

51. Lindgren, F., Rue, H. & Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *R. Stat. Soc.* **73**, 423–498 (2011).
52. Allison, P. D. Discrete-time methods for the analysis of event histories. *Sociol. Methodol.* **13**, 61 (1982).
53. Institute for Health Metrics and Evaluation. *Global Burden of Disease Study 2017 (GBD 2017) Socio-Demographic Index (SDI) 1950–2017*. http://ghdx.healthdata.org/record/ihme-data/gbd-2017-socio-demographic-index-sdi-1950%E2%80%932017 (GHDx, 2019).
54. Ahmad, O. B., Lopez, A. D. & Inoue, M. The decline in child mortality: a reappraisal. *Bull. World Health Organ.* **78**, 1175–1191 (2000).
55. Somoza, J. L. *Illustrative Analysis: Infant and Child Mortality in Colombia. World Fertility Survey Scientific Report No. 10* (International Statistical Institute, 1980).

56. Rutstein, S. O. *Infant and child mortality: levels, trends and demographic differentials. World Fertility Survey Scientific Report No. 43* (International Statistical Institute, 1984).
57. Burstein, R., Wang, H., Reiner, R. C. Jr & Hay, S. I. Development and validation of a new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories. *PLoS Med*. **15**, e1002687 (2018).
58. Graetz, N. et al. Mapping local variation in educational attainment across Africa. *Nature* **555**, 48–53 (2018).
59. Diggle, P. & Ribeiro, P. J. *Model-based Geostatistics* (Springer, 2007).
60. Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. M. TMB: automatic differentiation and Laplace approximation. *J. Stat. Softw*. **70**, 1–21 (2016).
61. Patil, A. P., Gething, P. W., Piel, F. B. & Hay, S. I. Bayesian geostatistics in health cartography: the perspective of malaria. *Trends Parasitol*. **27**, 246–253 (2011).
62. Phillips, D. E. et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Popul. Health Metr*. **12**, 14 (2014).
63. Schmertmann, C. P. & Gonzaga, M. R. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography* **55**, 1363–1388 (2018).
64. Walker, N., Hill, K. & Zhao, F. Child mortality estimation: methods used to adjust for bias due to AIDS in estimating trends in under-five mortality. *PLoS Med*. **9**, e1001298 (2012).
65. Silva, R. Child mortality estimation: consistency of under-five mortality rate estimates using full birth histories and summary birth histories. *PLoS Med*. **9**, e1001296 (2012).
66. Wilson, K. & Wakefield, J. Pointless spatial modeling. *Biostatistics* https://doi.org/10.1093/biostatistics/kxy041 (2018).
67. Wilson, K. & Wakefield, J. Child mortality estimation incorporating summary birth history data. Preprint at https://arxiv.org/abs/1810.04140 (2018).
68. WorldPop Dataset (WorldPop, accessed 25 July 2017); http://www.worldpop.org.uk/data/get_data/

**Acknowledgements** This work was primarily supported by grant OPP1132415 from the Bill & Melinda Gates Foundation.

**Author contributions** S.I.H. and R.B. conceived and planned the study. R.B., S.I.H., M.C., N.H., J.L, A.B., N.G. and S.W. identified and obtained data for this analysis. M.C., N.H., J.L, A.B. and S.W. extracted, processed and geo-positioned the data. R.B. and N.H. carried out the statistical analyses with assistance and input from S.I.H., A.O.-Z. and N.M. R.B., N.H., M.C., S.W., L.W., K.J. and L.E. prepared figures and tables. R.B. and L.B.M. wrote the first draft of the manuscript with assistance from S.I.H., S.P., L.E.S. and N.D.W. and all authors contributed to subsequent revisions. All authors provided intellectual input into aspects of this study.

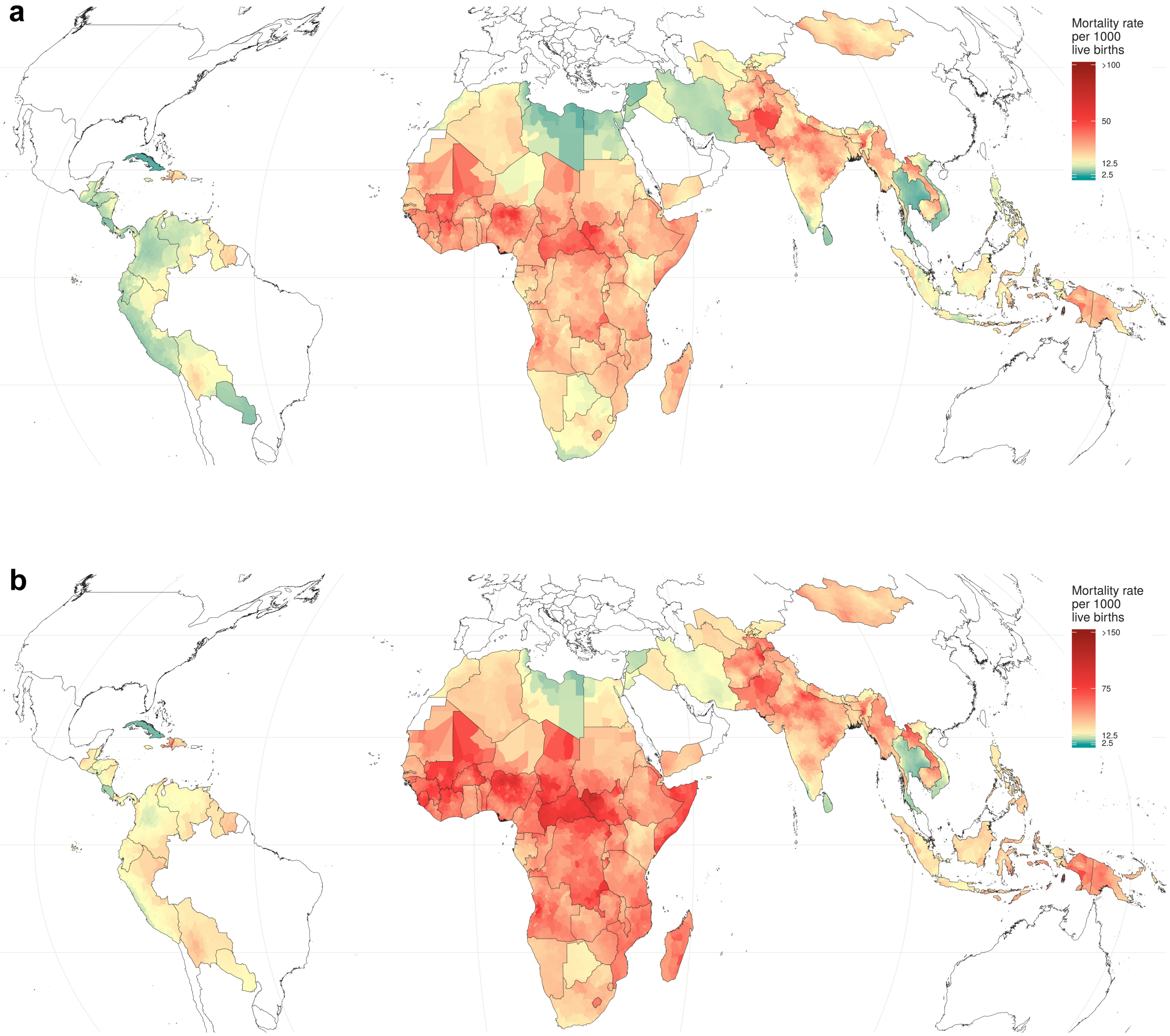
**Competing interests** This study was funded by the Bill & Melinda Gates Foundation. Co-authors employed by the Bill & Melinda Gates Foundation provided feedback on initial maps and drafts of this manuscript. Otherwise, the funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the final report, or decision to publish. The corresponding author had full access to all of the data in the study and had final responsibility for the decision to submit for publication.

**Additional information**

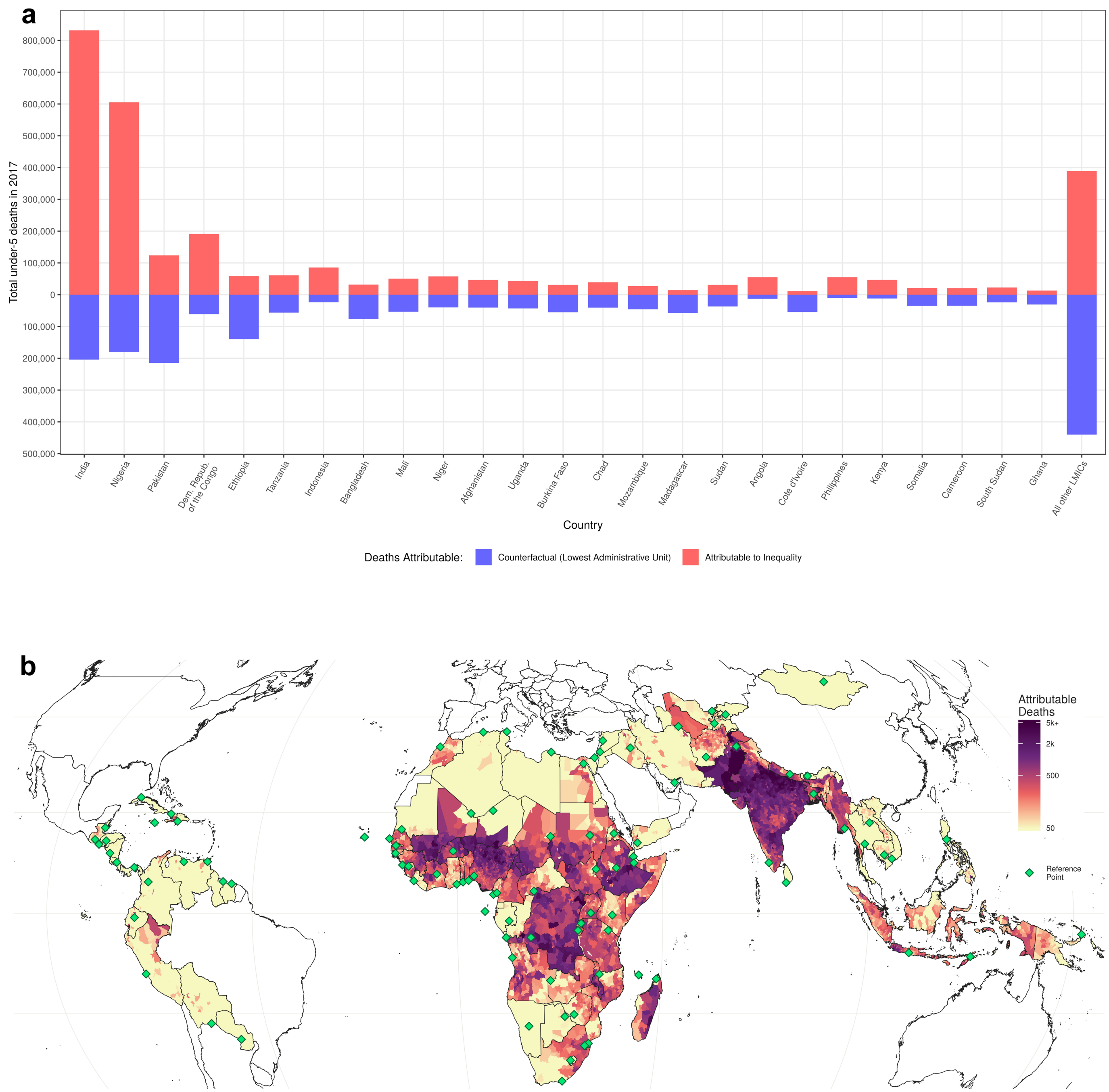
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1545-0.

**Correspondence and requests for materials** should be addressed to S.I.H.

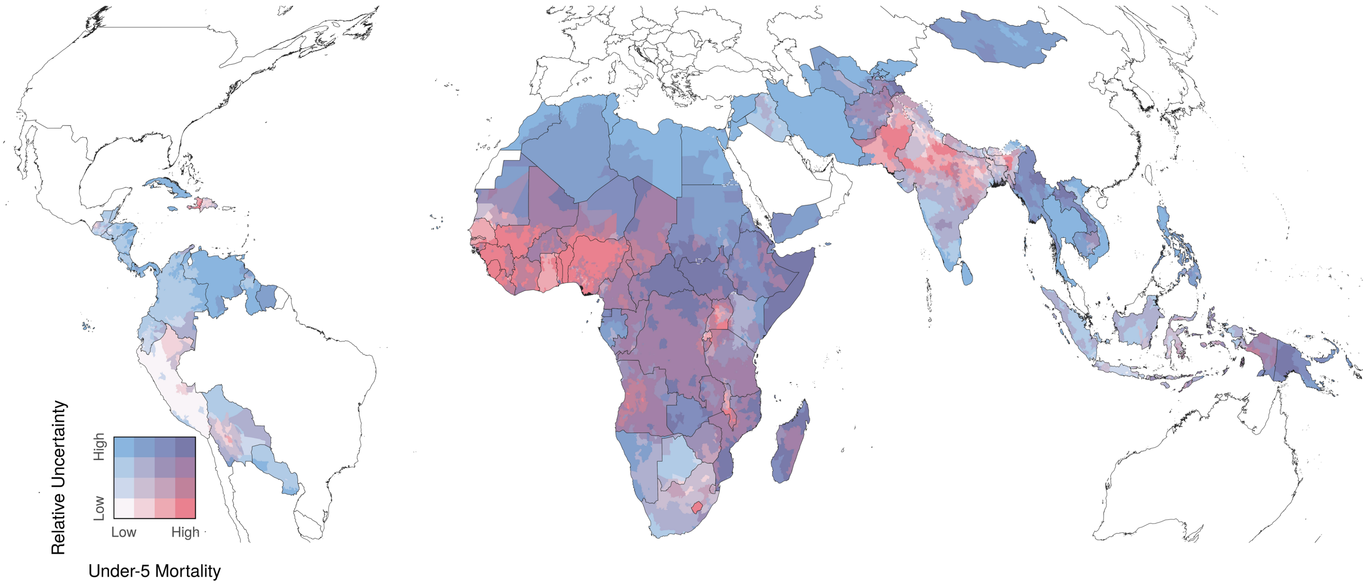
**Peer review information** *Nature* thanks Paloma Botella and the other, anonymous, reviewer(s) for their contribution to the peer review of this work



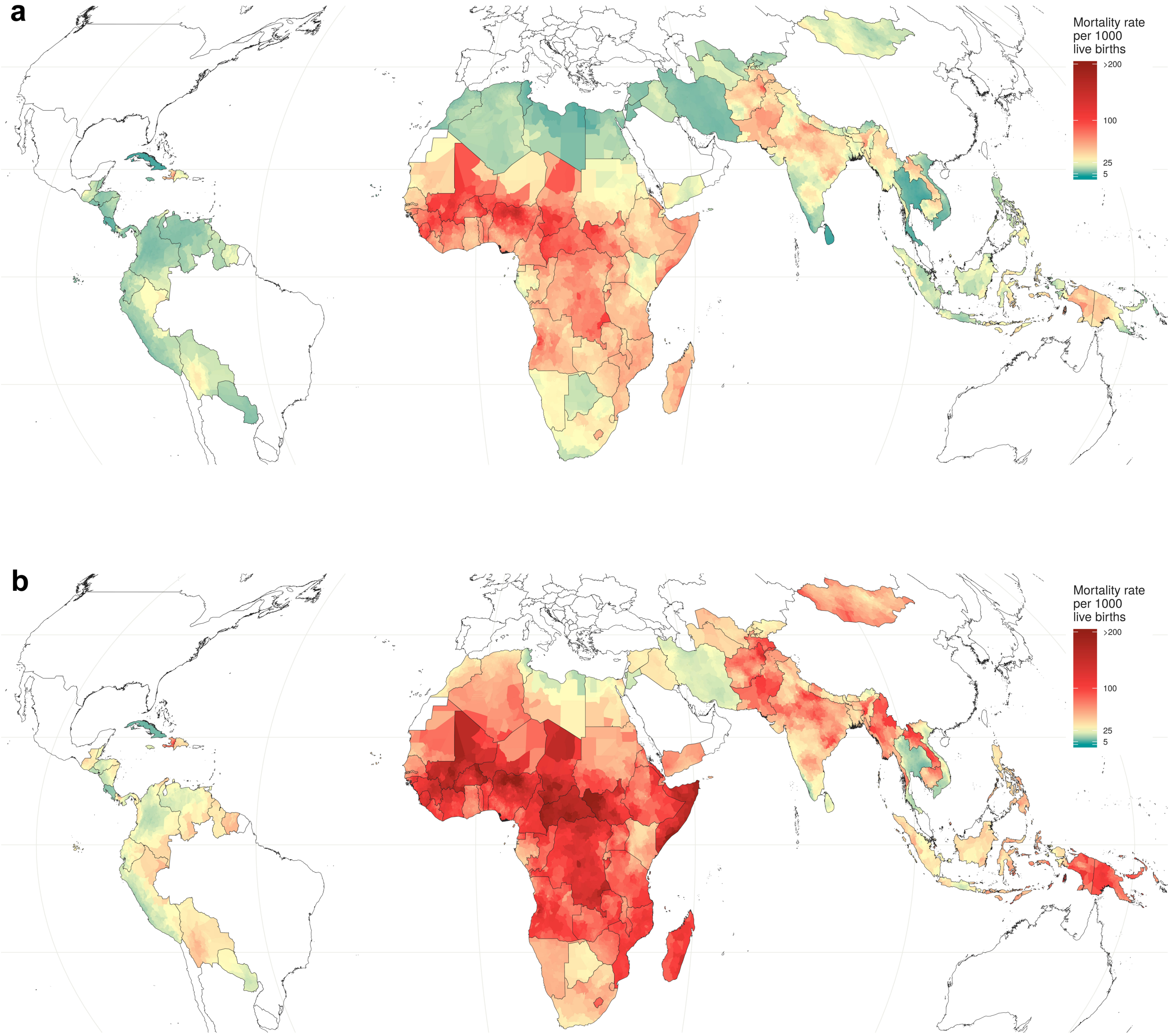
**Extended Data Fig. 1 | Neonatal and infant mortality rates in 2017.** **a**, **b**, Maps showing the mortality rates of neonates (**a**; birth to 28 days of age) and infants (**b**; under 1 year of age) across second administrative-level units in 2017. Note that the ranges in the keys are different for the two maps.



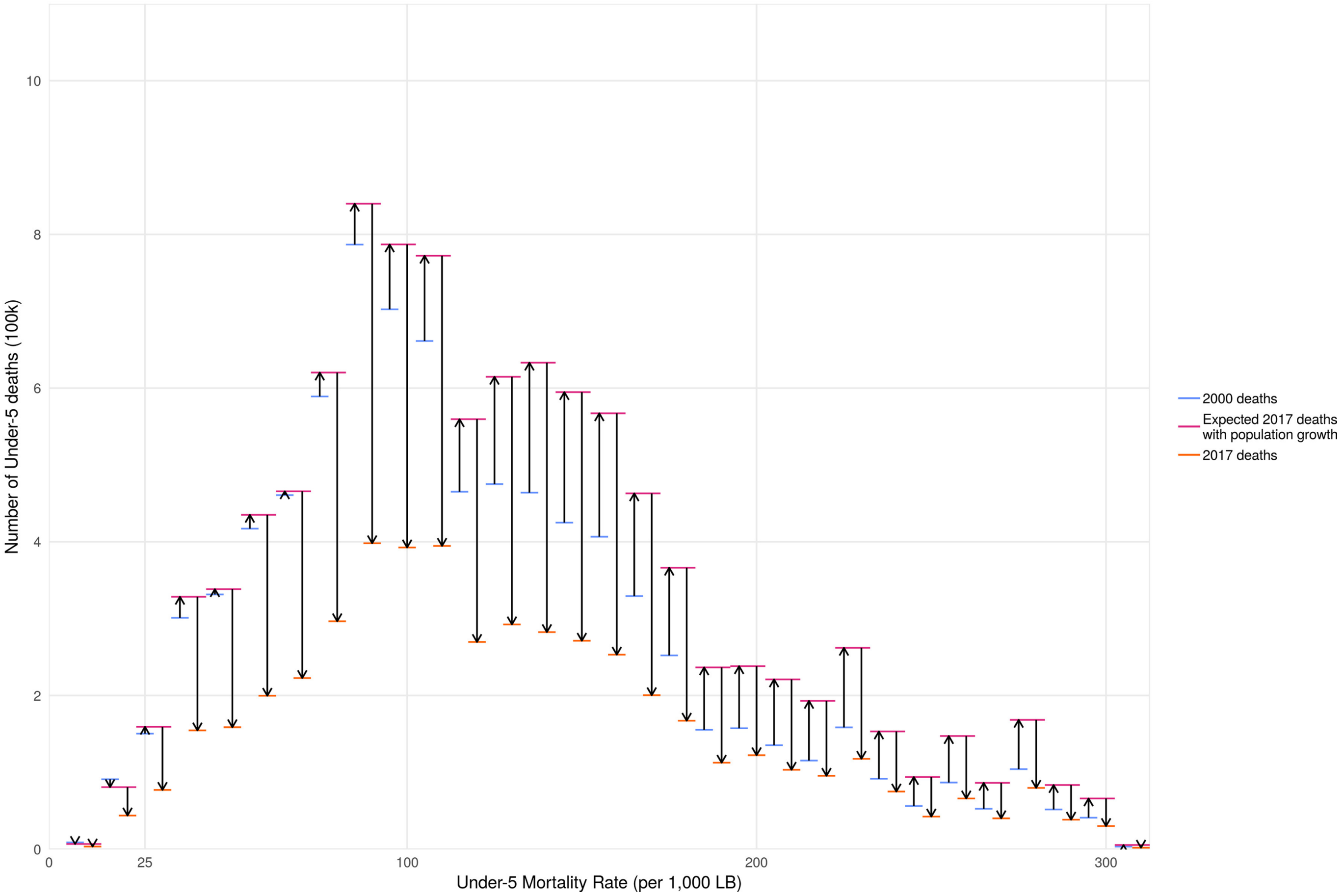
**Extended Data Fig. 2 | Impact of inequality on U5MR. a**, Potential reduction in the number of deaths that would have occurred if all second administrative-level units in the 20 countries with the greatest number of deaths of children under 5 in 2017 realized a homogenous U5MR that was equal to that of the lowest observed mortality rate in that country. In total, 66% of under-5 deaths could have been averted if all countries maintained mortality rates equal to the second administrative-level unit with lowest mortality. If this reference rate is set to the lowest observed rate across all of the 99 countries that were included in this study, 95% of under-5 deaths could have been averted. The size of each bar represents the total number of under-5 deaths in each country. The red portion of each bar indicates the number of deaths 'attributed' to geographical inequality in mortality rates, whereas the blue portion represents the number of deaths that would remain in the scenario in which all second administrative-level units within countries had the same mortality rate as the best-performing unit. **b**, Locations of under-5 deaths 'attributable' to geographical inequality, across all second administrative-level units in each country. Each country has one unit highlighted with a green diamond, which is the reference unit, or the location with the lowest mortality rate in the country in 2017.



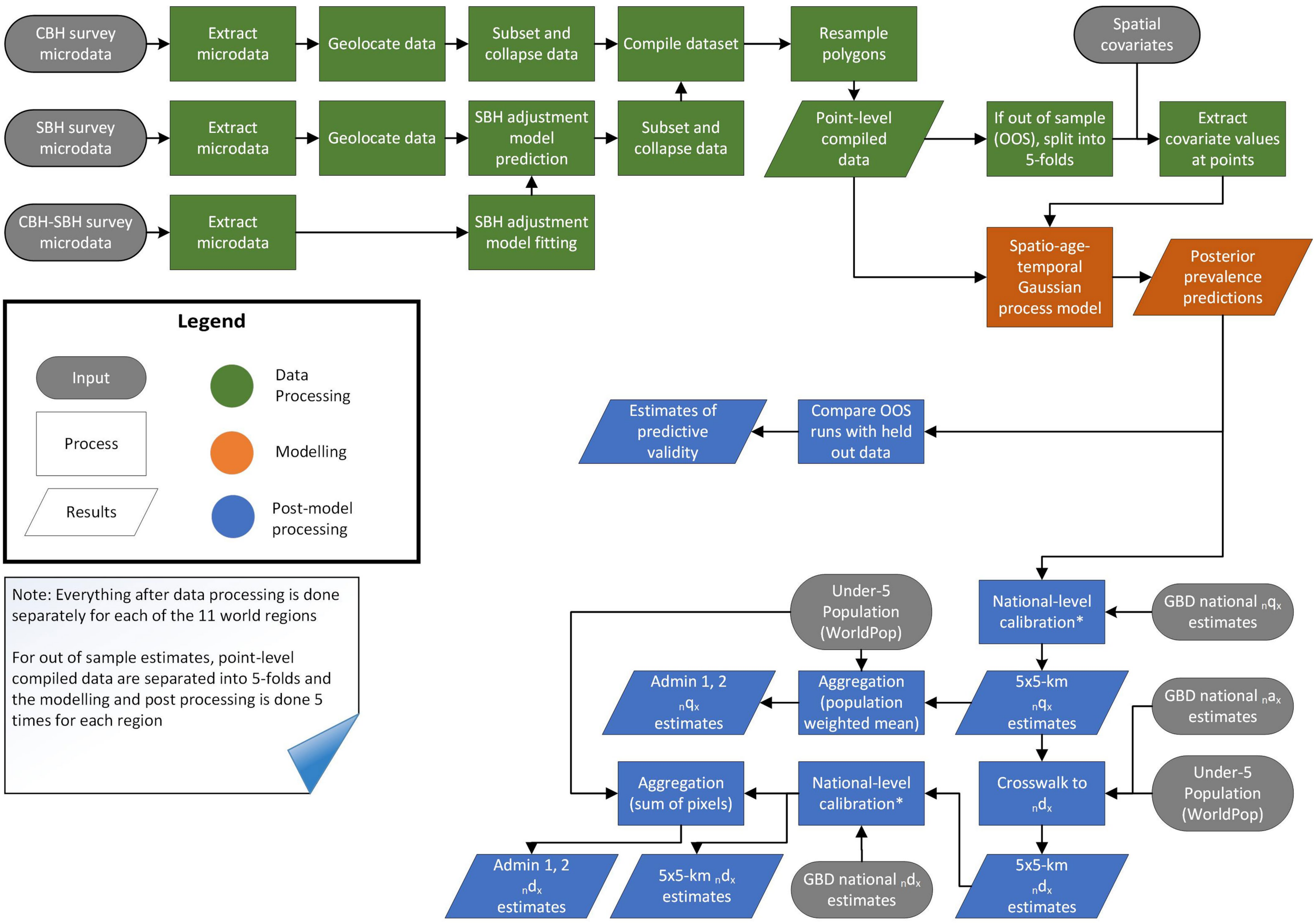
**Extended Data Fig. 3 | Relative uncertainty in U5MR estimates for 2017.** Relative uncertainty in second administrative-level estimates compared with mean estimated U5MRs in each second administrative-level unit for 2017. Mean rates and relative uncertainty are split into population-weighted quartiles. These cut-off points indicate the relative uncertainty minimum, 25th, 50th and 75th percentiles, and maximum, which are 0.29, 0.51, 0.63 and 0.86, and 3.12, respectively. The under-5 mortality minimum, 25th, 50th and 75th percentiles, and maximum are 1.4, 13.0, 22.9 and 44.8, and 190.6 deaths per 1,000 live births. Areas in which our estimates are more uncertain are coloured with a scale of increasing blue hue, whereas areas in which the mean estimates of U5MR are high are coloured with a scale of increasing red hue. Purple areas have high, but uncertain, estimates of U5MRs. White areas have low relative mortality, with fairly certain estimates. Relative uncertainty is defined as the ratio of the width of the 95% uncertainty interval to the mean estimate.



**Extended Data Fig. 4 | Lower and upper uncertainty interval boundaries for U5MR mortality estimates in 2017. a**, **b**, Lower (**a**) and upper (**b**) 95% uncertainty intervals for U5MR estimates across the second administrative-level units in 99 countries.



**Extended Data Fig. 5 | The counteracting forces of population change and mortality rate decline on total number of under-5 deaths.** Arrow plots show the mortality rate strata (bins of 10 per 1,000 livebirths) in 2000.



**Extended Data Fig. 6 | Flowchart summarizing analytical process.** Standard demographic notation were used. *n*, length of age bin; *x*, starting age of age bin; *d*, number of deaths; *q*, probability of death; *a*, average time lived in age bin by those who died in the age bin.

natureresearch

Corresponding author(s): Simon Hay
Last updated by author(s): Jul 9, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No primary data collection was carried out for this analysis. |
| Data analysis | This analysis was carried out using R version 3.5.0. The main statistical model used the Template Model Builder (TMB) software version 1.7.14 in R. All code used for these analyses is publicly available online at https://github.com/ihmeuw/lbd/tree/u5mr-2019. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The findings of this study are supported by data available in public online repositories, data publicly available upon request of the data provider, and data not publicly available due to restrictions by the data provider. Non-publicly available data were used under license for the current study but may be available from the authors upon reasonable request and with permission of the data provider. A detailed table of data sources and availability can be found in Supplementary Table 8.1. The full output of the analyses are publicly available in the Global Health Data Exchange (http://ghdx.healthdata.org/) and can further be explored via data visualization tools (https://healthdata.org/lbd/under5).

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | This observational study incorporated all available survey data sources that met the inclusion criteria as described in the Methods section. The combined dataset used in this analysis contained 25.1 million births and 1.3 million child deaths. |
| Data exclusions | Surveys were excluded due to missingness greater than 10% in date of birth or death and children ever born or died, unrealistic geographic trends compared to other surveys in nearby country-years, inability to match the microdata to geographic locations, or non-standard methodology. A full list of excluded surveys is included in Supplementary Table 8.2. |
| Replication | This is an observational study using many years of survey and surveillance data and could be replicated. |
| Randomization | This analysis is an observational mapping study, there were no experimental groups. |
| Blinding | Blinding was not relevant to this study, as it was an observational study using survey and surveillance data. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

https://doi.org/10.1038/s41586-019-1647-8

# Regulation of lifespan by neural excitation and REST

Joseph M. Zullo[1,6], Derek Drake[1,6], Liviu Aron[1], Patrick O'Hern[1], Sameer C. Dhamne[2], Noah Davidsohn[1], Chai-An Mao[3], William H. Klein[4], Alexander Rotenberg[2], David A. Bennett[5], George M. Church[1], Monica P. Colaiácovo[1] & Bruce A. Yankner[1]*

**The mechanisms that extend lifespan in humans are poorly understood. Here we show that extended longevity in humans is associated with a distinct transcriptome signature in the cerebral cortex that is characterized by downregulation of genes related to neural excitation and synaptic function. In *Caenorhabditis elegans*, neural excitation increases with age and inhibition of excitation globally, or in glutamatergic or cholinergic neurons, increases longevity. Furthermore, longevity is dynamically regulated by the excitatory–inhibitory balance of neural circuits. The transcription factor REST is upregulated in humans with extended longevity and represses excitation-related genes. Notably, REST-deficient mice exhibit increased cortical activity and neuronal excitability during ageing. Similarly, loss-of-function mutations in the *C. elegans* REST orthologue genes *spr-3* and *spr-4* elevate neural excitation and reduce the lifespan of long-lived *daf-2* mutants. In wild-type worms, overexpression of *spr-4* suppresses excitation and extends lifespan. REST, SPR-3, SPR-4 and reduced excitation activate the longevity-associated transcription factors FOXO1 and DAF-16 in mammals and worms, respectively. These findings reveal a conserved mechanism of ageing that is mediated by neural circuit activity and regulated by REST.**

Studies in invertebrate and mammalian models suggest that the nervous system plays a role in the regulation of ageing[1,2]. In the nematode *C. elegans*, ablation of specific sensory or neurosecretory neurons alters lifespan[3–6], and lifespan extension from reduced insulin/IGF-like signalling can be reversed by restoring function specifically in neurons[7]. However, whether the activity state of the nervous system affects the ageing process is unclear. Here we describe a conserved mechanism of ageing that is mediated by global neural activity and regulated by REST.

## Neural excitation and longevity

Previous studies of the ageing human brain have shown dynamic changes in gene expression that distinguish young adults from the ageing population[8]. However, the recent expansion of data from ageing human cohorts has enabled partitioning of the ageing population into subgroups based on transcriptome profiling. To gain insight into changes in gene expression in the brain that are associated with extended human longevity, we analysed RNA sequencing (RNA-seq)[9,10] and microarray data[11] from the frontal cortex of aged individuals with intact cognitive function in three cohorts: ROSMAP, CommonMind Consortium (CMC), and Gibbs. We compared age-associated genes between all individuals and derived Pearson correlation coefficients. Hierarchical clustering suggested that the ageing population partitioned into three groups (Extended Data Fig. 1a–c). The most significant changes associated with the extended longevity groups ($\geq$85 versus $\leq$80 years of age) were downregulation of genes related to neural excitation and synaptic function, and upregulation of genes involved in immune function (Fig. 1a–d, Extended Data Fig. 1d–f, Supplementary Tables 1–6). This did not change after adjusting for amyloid deposits and neurofibrillary tangles in the ROSMAP cohort (data not shown). Meta-analysis of gene ontology (GO) enrichment in each cohort indicated that terms related to excitatory, but not inhibitory, synaptic transmission were enriched in the downregulated genes (Extended Data Fig. 1g, Supplementary Table 7). These results suggest that extended human longevity may be associated with reduced excitatory neurotransmission.

To explore the neural regulation of longevity, we used *C. elegans*, a well-established model system of ageing. We monitored neural excitation in *C. elegans* by GCaMP calcium imaging in the glutamatergic ASH neurons[12]. In wild-type worms, we observed rapid, transient pulses of GCaMP fluorescence indicative of neuronal excitation (Supplementary Video 1). Calcium influx in ASH neurons increased during normal ageing from adult day 1–2 to day 12–16 (Fig. 1e). To determine the effect of decreasing calcium influx on lifespan, worms were treated with nemadipine, an inhibitor of L-type calcium channels that reduces neural excitation (Fig. 1f). Continuous treatment with nemadipine beginning at adult day 1 extended lifespan (Fig. 1g). Moreover, incubation of worms with ivermectin, an agonist of invertebrate glutamate-gated chloride channels, suppressed neural excitation and resulted in a dose-dependent extension of mean lifespan (Fig. 1h, i). Nemadipine and ivermectin also extended lifespan when administered at day 8, when feeding activity has largely abated (Extended Data Fig. 2a), suggesting that the drugs do not act through caloric restriction. Furthermore, worm motility was preserved (Extended Data Fig. 2b, Supplementary Videos 4–6). These results suggest that global inhibition of neural excitation extends lifespan in *C. elegans*.

To explore the neural systems that mediate lifespan, we expressed a transgenic *Drosophila* histamine-gated chloride channel (HisCl1) in different neuronal populations in *C. elegans*[13]. Addition of histamine, which is not endogenously produced by worms, activates HisCl1 and inhibits neural excitation. First, we expressed HisCl1 under the control of a pan-neuronal promoter. Continuous incubation with histamine beginning on adult day 1 or day 8 extended mean lifespan (Extended Data Fig. 3a, b, i). Histamine had no effect on the lifespan of wild-type worms that did not express HisCl1 (Supplementary Table 22).

---

[1]Department of Genetics, Harvard Medical School, Boston, MA, USA. [2]F. M. Kirby Neurobiology Center, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [3]Department of Ophthalmology and Visual Science, The University of Texas McGovern Medical School, Houston, TX, USA. [4]Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [5]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [6]These authors contributed equally: Joseph Zullo, Derek Drake. *e-mail: bruce_yankner@hms.harvard.edu

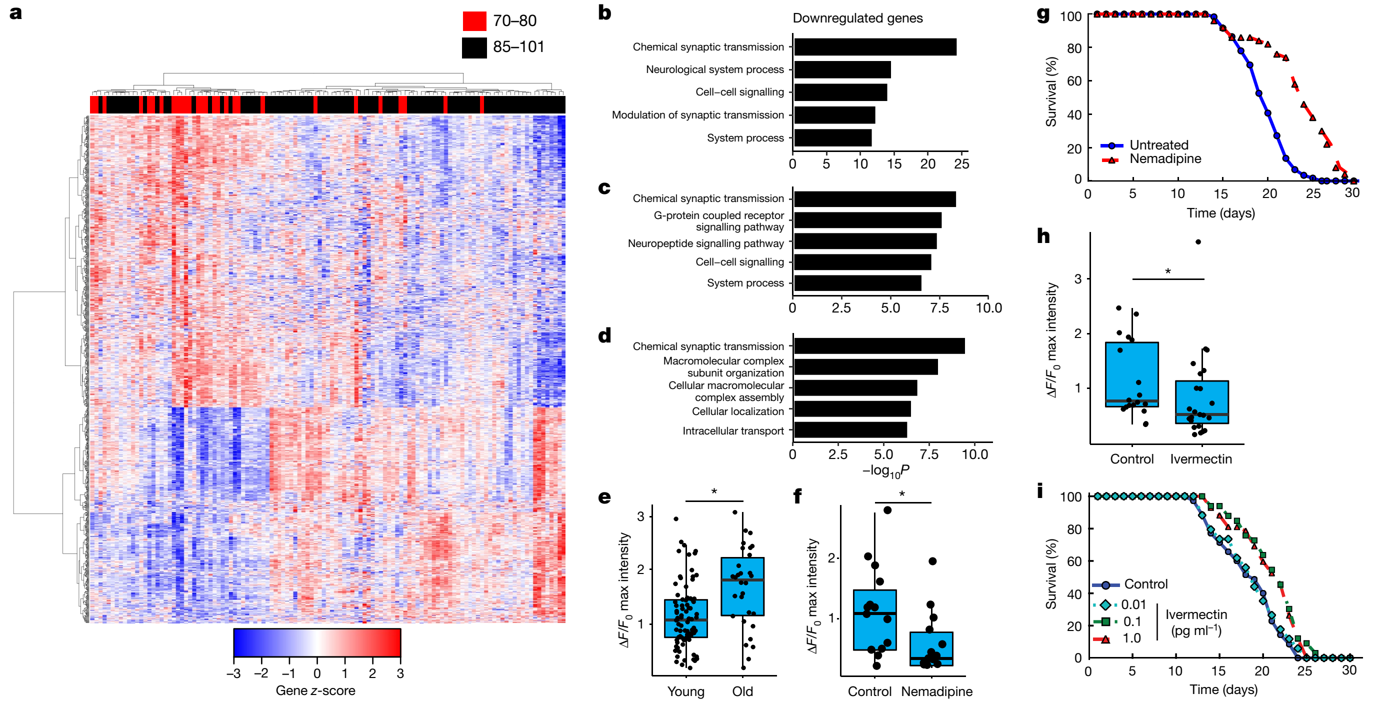


**Fig. 1 | Neural excitation and longevity in humans and *C. elegans*.** **a**, Analysis of the cortical transcriptome profile in cognitively intact aged individuals from the ROSMAP cohort. Unsupervised hierarchical clustering shows a transcriptional signature of downregulated and upregulated genes associated with extended longevity. **b**–**d**, Most significantly enriched gene ontology (GO) terms for downregulated genes associated with extended longevity (≥85 versus ≤ 80 years of age) in the ROSMAP (**b**; dorsolateral prefrontal cortex, $n = 117$), CMC (**c**; dorsolateral prefrontal cortex, $n = 155$), and Gibbs (**d**; frontal cortex, $n = 37$) cohorts. $P$ values calculated by Fisher's exact test (see Methods). **e**, Ageing *C. elegans* exhibit increased neuronal excitation. Maximum GCaMP fluorescence intensity changes in ASH neurons of young adult (days 1–2) and older (days 12–16) worms. Young, $n = 82$ worms; old, $n = 30$ worms. $^*P = 3.6 \times 10^{-4}$ by Mann–Whitney $U$-test. In box plots: centre line, median; lower and upper hinges, first and third quartiles; upper whisker, hinge to the largest value no further than 1.5 × the interquartile range (IQR) from the hinge; lower whisker, hinge to the smallest value at most 1.5 × IQR of the hinge. **f**, The L-type calcium channel blocker nemadipine (2 μM) represses neural excitation. Control, $n = 14$; nemadipine, $n = 13$. $^*P = 0.029$, Mann–Whitney $U$-test. **g**, Nemadipine extends lifespan. Worms were continuously treated with 2 μM nemadipine beginning at adult day 1. $P = 7.7 \times 10^{-11}$, log-rank test. Control, $n = 59$; nemadipine, $n = 50$, replicated three times. **h**, The chloride channel agonist ivermectin (1 pg ml$^{-1}$) reduces neural excitation. Control, $n = 18$; ivermectin, $n = 23$. $^*P = 0.038$, Mann–Whitney $U$-test. **i**, Extension of lifespan by continuous treatment with ivermectin beginning at adult day 1 (control, $n = 35$; 0.01 pg ml$^{-1}$, $n = 34$, $P = 0.62$ ; 0.1 pg ml$^{-1}$, $n = 33$, $P = 1.5 \times 10^{-3}$; 1 pg ml$^{-1}$, $n = 42$, $P = 1.9 \times 10^{-3}$, log-rank test), replicated three times. Summary statistics for all individual lifespan experiments are in Supplementary Table 22.

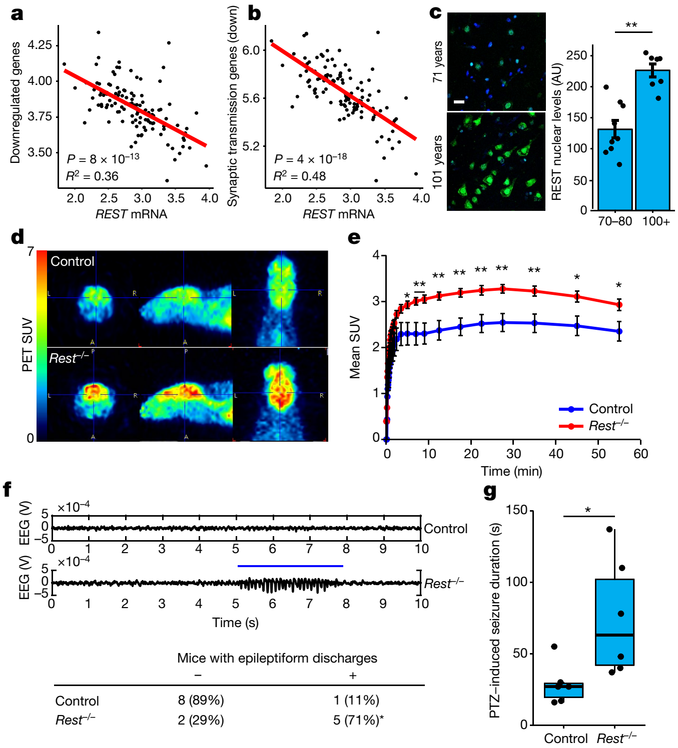
We next expressed the HisCl1 channel in glutamatergic and cholinergic neurons, the major excitatory neuronal populations in *C. elegans*. Repression of excitation in either population robustly extended lifespan whether initiated at adult day 1 or day 8 (Extended Data Fig. 3c–f, i). Expression of HisCl1 in γ-aminobutyric acid (GABA)ergic neurons using an *unc-47* driver extended lifespan when initiated at day 1, but reduced lifespan when initiated at day 8 (Extended Data Fig. 3g–i). GCaMP imaging showed that addition of histamine to *unc-47*:HisCl1 worms at day 1 resulted in marked and persistent suppression of excitation in ASH neurons, which was not observed after addition at day 8 (Extended Data Fig. 3j). Thus, blockade of GABAergic neurotransmission early in adult life may result in compensatory downregulation of excitation in other neuronal populations. Together, these results suggest that continuous or late-life repression of neural excitation in multiple neuronal cell populations extends lifespan in *C. elegans*.

To explore the effect of hyperexcitation on lifespan, we suppressed GABAergic neurotransmission by using RNA-mediated inhibition (RNAi), which would be predicted to be less extensive than histamine/HisCl1-mediated blockade. When worms were treated with RNAi against the GABA vesicular transporter *unc-47*, there was a robust increase in excitation in ASH neurons and a reduction in lifespan (Extended Data Fig. 4a, b). Thus, the effects of neurotransmission on lifespan are bidirectional; lifespan is extended by reducing excitation and shortened by increasing excitation.

Neural activity can regulate neuropeptide secretion. To investigate the role of neuropeptide signalling in lifespan regulation, we evaluated worms in which the function of the EGL-3 proprotein convertase was blocked by an *egl-3* mutation or *egl-3* RNAi. The *egl-3* mutant exhibited robust lifespan extension (Extended Data Fig. 4c). Lifespan was also extended in worms treated with *egl-3* RNAi (Extended Data Fig. 4d), consistent with previous results[14]. Similar extension of lifespan was observed in the glutamatergic loss-of-function *eat-4* mutant and the synaptic transmission *unc-13* mutant (Extended Data Fig. 4c). These results suggest that both synaptic neurotransmission and peptidergic signalling contribute to the regulation of lifespan.

## REST modulates excitation in the ageing brain

We have previously demonstrated that the transcriptional repressor REST is induced in the ageing brain[15]. Genes that were downregulated in the brain in individuals with extended longevity were enriched for the canonical REST RE1 motif in all three ageing cohorts (ROSMAP, $P = 5 \times 10^{-12}$; CMC, $P = 8 \times 10^{-4}$; Gibbs, $P = 1 \times 10^{-2}$; Supplementary Tables 1, 3, 5). Moreover, unbiased analysis by chromatin immunoprecipitation and sequencing (ChIP–seq) showed that REST was the most strongly implicated transcription factor in multiple ENCODE datasets (Supplementary Tables 8–11). Furthermore, the downregulated gene set was highly enriched for neuronal REST target genes (Supplementary Table 12). Expression of these downregulated genes, as well as an index of synaptic gene expression, were inversely related to expression of *REST* mRNA (Fig. 2a, b, Extended Data Fig. 5a, b). Furthermore, levels of REST in nuclei were elevated in prefrontal cortical neurons in centenarians relative to individuals who



| | Mice with epileptiform discharges − | Mice with epileptiform discharges + |
|---|---|---|
| Control | 8 (89%) | 1 (11%) |
| *Rest*$^{-/-}$ | 2 (29%) | 5 (71%)* |

**Fig. 2 | REST regulates neural excitation in the ageing brain and is associated with extended longevity. a, b**, Expression of genes downregulated in individuals with extended longevity (≥85 versus ≤80 years old) is inversely related to levels of *REST* mRNA. Linear regression analysis of the mean expression of all downregulated genes (**a**) and downregulated genes associated with the synaptic transmission GO term (**b**). Data are from the ROSMAP cohort. Each point represents an individual case ($n = 117$). $P$ values derived by $t$-tests of the regression line slopes. **c**, Increased nuclear REST levels in the prefrontal cortex of centenarians. Left, immunofluorescence labelling for REST (green, rabbit polyclonal; Bethyl laboratories) and DAPI (blue) in human prefrontal cortex. Scale bar, 40 μm. Right, nuclear REST levels in cognitively intact individuals 70–80 years ($n = 9$) and >100 years ($n = 7$) of age. AU, arbitrary units. Mean ± s.e.m., $^{**}P = 1.5 \times 10^{-4}$, Student's $t$-test. **d**, REST represses neural excitation in the mouse cerebral cortex. Shown are images from PET–CT scanning of fluorodeoxyglucose ($^{18}$F-FDG) uptake in 18-month-old *Nestin-Cre;Rest*$^{lx/lx}$ (*Rest*$^{-/-}$) and age-matched *Rest*$^{lx/lx}$ (control) mice. **e**, Mean ± s.e.m. standardized uptake value (SUV) at increasing time intervals after injection of $^{18}$F-FDG. $n = 7$ mice per group; $^{*}P < 0.05$, $^{**}P < 0.01$; Mann–Whitney $U$-test. **f**, Increased epileptiform discharges in aged REST-deficient mice. Top, EEG recording from *Rest*$^{-/-}$ and age-matched control mice. Bottom, number of mice with at least one epileptiform discharge (≥3 s) in a 48-h recording. Control, $n = 9$; *Rest*$^{-/-}$, $n = 7$. $^{*}P = 0.035$, Fisher's exact test. **g**, Seizure duration after administration of PTZ (40 mg kg$^{-1}$). Control, $n = 6$; *Rest*$^{-/-}$, $n = 6$ mice. $^{*}P = 0.016$, Mann–Whitney $U$-test.

were 70–80 years of age (Fig. 2c). Although expression of *REST* mRNA is upregulated in the brain during ageing[15], increased *REST* mRNA expression did not distinguish between the extended longevity and normal ageing groups based on RNA-seq (Supplementary Tables 1, 3). However, for a given level of expression of *REST* mRNA, there is greater gene downregulation in the extended longevity group (Extended Data Fig. 5c, d). These results suggest that REST repressor function is upregulated in the brain in individuals with extended longevity, resulting in downregulation of genes that mediate excitation and synaptic function.

To assess the role of REST as a modulator of neural activity in the ageing mammalian brain, we examined the uptake of fluorodeoxyglucose (FDG) by positron emission tomography and computerized tomography (PET–CT) in the brains of 18-month-old *Nestin-Cre:Rest*$^{lx/lx}$ (*REST* conditional knockout (cKO)) mice and littermate controls[15] (Extended Data Fig. 5e). *REST* cKO mice showed elevated cortical $^{18}$F-FDG uptake, indicative of increased neural activity (Fig. 2d, e). Previous studies suggest that REST can modulate excitability in mouse models of epilepsy induced by kindling or kainate[16,17]. To assess excitability during ageing, we performed electroencephalographic (EEG) recordings of *REST* cKO mice (Supplementary Table 13). Intermittent epileptiform discharges were more frequent in 22.5- to 23-month-old *REST* cKO mice than in controls (Fig. 2f). Furthermore, challenge with the GABA antagonist pentylenetetrazol (PTZ) increased seizure duration in *REST* cKO mice relative to controls (Fig. 2g), with a trend towards increased mortality (Extended Data Fig. 5f). These results suggest that REST globally represses neural activity and prevents hyperexcitation in the ageing brain.

## *C. elegans* REST orthologues regulate longevity

The *C. elegans* gene *spr-4* encodes a structural and functional orthologue of mammalian REST that protects against toxic stressors, such as reactive oxygen species and amyloid-β protein[15]. To determine whether *spr-4* modulates lifespan, we induced endogenous expression of *spr-4* by using the RNA-guided endonuclease Cas9 as a programmable transcription factor[18]. A nuclease-deficient variant of Cas9 (dCas9) was fused to the transcriptional activator VP16 (dCas9::VP64) and stably introduced into *C. elegans* together with four small guide RNAs (sgRNAs) targeting the *spr-4* promoter. This resulted in a modest elevation in *spr-4* mRNA and protein expression, and a significant increase in mean lifespan (Extended Data Fig. 6a–d). Expression of dCas9::VP64 and *spr-4* sgRNAs in worms with the loss-of-function allele *spr-4* (*tm465*) did not affect lifespan, suggesting specificity for *spr-4* (Extended Data Fig. 6e). Moreover, overexpression of *spr-4* robustly reduced excitation in ASH neurons (Extended Data Fig. 6f). Thus, SPR-4 both represses neural excitation and extends lifespan.

The forkhead transcription factor DAF-16 is the central downstream target of the DAF-2–insulin/IGF-like signalling pathway that regulates lifespan in *C. elegans*. RNAi-mediated knockdown of *daf-16* prevented extension of lifespan by overexpression of *spr-4* (Extended Data Fig. 7a). Furthermore, extension of lifespan by the neural excitation inhibitors nemadipine and ivermectin was also dependent on *daf-16*, and ivermectin elevates both total and nuclear levels of DAF-16 (Extended Data Fig. 7b–f). Thus, DAF-16 mediates extension of lifespan by *spr-4* and neural suppression.

To further explore the effects of REST orthologues on the DAF-2–DAF-16 signalling pathway, we performed *daf-2* RNAi in wild-type worms, *spr-3* and *spr-4* loss-of-function mutants, and an *spr-4;spr-3* double mutant. As previously shown[19], *daf-2* RNAi extends lifespan by about 50% in wild-type worms. Mutations in *spr-3*, *spr-4*, or both reduced the extension of lifespan by *daf-2* RNAi (Extended Data Fig. 8a). Mutations in *spr-3* and *spr-4* also reduced the lifespan extension associated with the loss-of-function *daf-2*(*e1370*) allele; the greatest reduction occurred in the *spr-4;spr-3;daf-2* triple mutant (Fig. 3a). The *spr-3* and *spr-4* mutations did not affect lifespan in a wild-type background (Extended Data Fig. 8b), in contrast to a previous report that suggested that *spr-3* mutations altered lifespan[20]. These results suggest that *spr-3* and *spr-4* contribute to the regulation of lifespan by the insulin/IGF-like signalling pathway in worms.

We next investigated whether SPR-3 and SPR-4 function in neurons to regulate lifespan. To address this question, we used a *C. elegans* line in which RNAi was abolished by deletion of the double-stranded RNA transporter SID-1, but restored specifically in neurons by a *sid-1* transgene driven by a neuron-specific promoter[21]. These alleles were crossed into the *daf-2(e1370)* mutant background, and RNAi-mediated knockdown of both *spr-3* and *spr-4* was performed. Neuron-targeted knockdown of *spr-3* and *spr-4* significantly reduced lifespan (Fig. 3b, Extended Data Fig. 8c). Thus, neuronal expression of *spr-3* and *spr-4* contributes to lifespan extension in the *daf-2* mutant.

## SPR-3 and SPR-4 regulate neural excitation

To gain further insight into the effects of SPR-3 and SPR-4 on DAF-2 function and lifespan, we performed RNA-seq on wild-type worms and

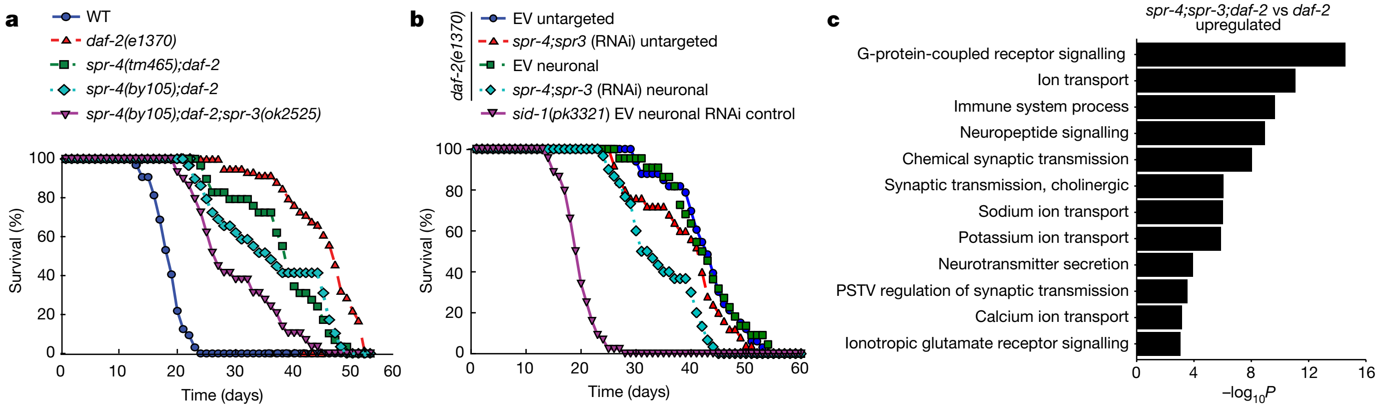


**Fig. 3 | *C. elegans* REST orthologues mediate longevity in *daf-2* loss-of-function mutants. a**, The REST orthologues *spr-4* and *spr-3* are required for maximal longevity in *daf-2* mutant worms. Lifespan analysis was performed on wild-type (WT) and *daf-2(1370)* loss-of-function mutant worms, and the indicated combinations of *daf-2* and *spr-4;spr-3* mutations. The *spr-4;spr-3* mutations significantly reduced the lifespan of *daf-2* mutant worms. $n = 29$–$59$ worms per genotype, replicated at least three times per genotype. $P < 0.001$ for all curves relative to *daf-2*, by log-rank test. **b**, Neuronal expression of *spr-3* and *spr-4* mediate lifespan extension in *daf-2* mutant worms. Lifespans of worms with neuronal targeting of RNAi by neuronal expression of a *sid-1* transgene in otherwise *sid-1* null *daf-2(1370)* mutants, or untargeted RNAi in *sid-1* wild-type *daf-2(1370)* mutants. Lifespan effect of neuronal targeting of *spr-4;spr-3* RNAi versus empty vector (EV) control RNAi is significant by log-rank test ($P = 2.5 \times 10^{-6}$), $n = 22$–$56$ worms per curve replicated at least four times. **c**, SPR-3 and SPR-4 repress genes that mediate neural excitation. Shown are significantly enriched GO terms for upregulated genes related to neural excitation in RNA-seq analysis of day 2 *spr-4;spr-3;daf-2* triple mutants versus *daf-2* single mutant worms. *P* values calculated using Fisher's exact test (see Methods), $n = 3$ biological replicates per genotype.

worms with mutations in *spr-4* and *spr-3*, *daf-2*, or *spr-4*, *spr-3* and *daf-2* (triple mutants) (Extended Data Fig. 8d, Supplementary Table 14). The comparison of triple mutant *spr-4;spr-3;daf-2* worms with single mutant *daf-2* worms was notable for highly significant changes in the transcriptome that were enriched for GO terms related to neural excitation, signalling and synaptic function (Fig. 3c, Supplementary Table 15). Furthermore, gene expression changes in *daf-2* mutants and *spr-4;spr-3* mutants overlapped (Extended Data Fig. 8e), and genes that were downregulated in *daf-2* mutants but upregulated in the *spr-4;spr-3;daf-2* triple mutants were enriched for GO terms related to neural excitation (Supplementary Table 16). These results suggest that repression of neuronal genes is a conserved regulatory feature of REST and its worm orthologues.

A central question is whether SPR-3 and SPR-4 affect lifespan by suppressing neural excitation. GCaMP calcium imaging showed that neural excitation was strongly suppressed in *daf-2* mutants, both in young adult worms and during ageing (Fig. 4a, b, Extended Data Fig. 8f, Supplementary Videos 1–3). The *spr-4;spr-3* mutations partially restored neural excitation in *daf-2* mutants (Fig. 4a, b), but did not increase excitation in wild-type worms (Extended Data Fig. 8g). Suppression of excitation in *daf-2* mutants was not mediated by neuronal DAF-16 (Extended Data Fig. 8h). However, inhibition of neural excitation with ivermectin reversed the lifespan-shortening effect of *spr-4;spr-3* mutations in *daf-2* mutant worms (Fig. 4c). Thus, SPR-3 and SPR-4 contribute to the extreme longevity of *daf-2* mutant worms by repressing neural excitation.

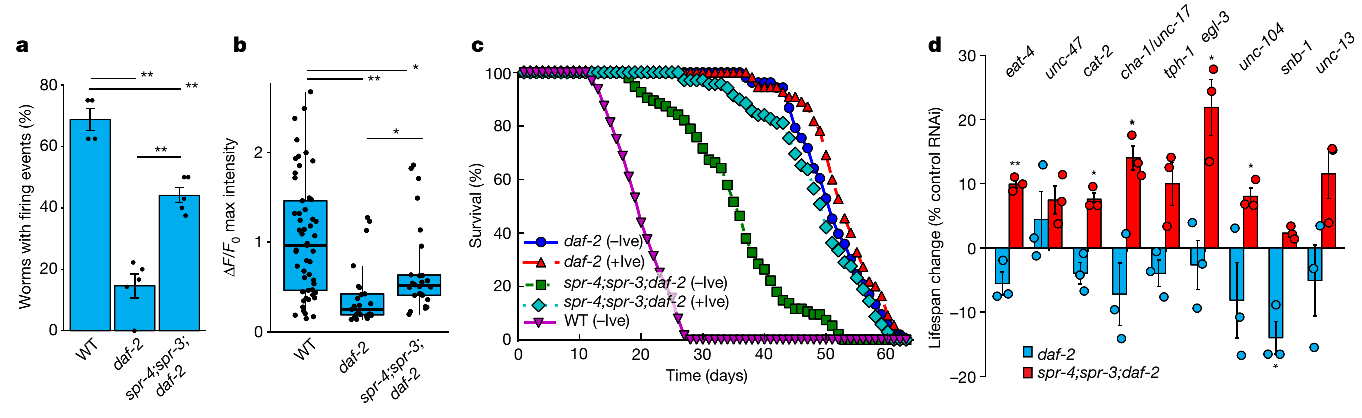


**Fig. 4 | SPR-3 and SPR-4 suppress multiple neurotransmitter and neuropeptide systems to extend lifespan in *daf-2* mutant worms. a**, Neural excitation is suppressed in *daf-2* mutants and partially restored by *spr-4;spr-3* mutations. GCaMP imaging was performed in ASH neurons. Shown is the fraction of worms with at least one firing event in a 2-min recording. Mean ± s.e.m., $n = 4$–$5$ independent experiments. $^{**}P = 7.9 \times 10^{-7}$ (*daf-2* versus wild-type), $P = 1.5 \times 10^{-4}$ (*spr-4;spr-3;daf-2* versus *daf-2*); $P = 0.0011$ (*spr-4;spr-3;daf-2* versus wild-type); ANOVA with post hoc Tukey test. **b**, Quantification of GCaMP fluorescence changes in day 2 worms: wild-type, $n = 53$; *daf-2*, $n = 25$; *spr-4;spr-3;daf-2*, $n = 26$. $^{**}P = 3.1 \times 10^{-6}$ (*daf-2* versus wild-type), $P = 1.5 \times 10^{-3}$ (*spr-4;spr-3;daf-2* versus *daf-2*), $^{*}P = 0.018$ (*spr-4;spr-3;daf-2* versus wild-type); Mann–Whitney *U*-test with multiple testing correction by Holm's method. **c**, Inhibition of neural excitation by ivermectin (+Ive, 10 pg ml$^{-1}$) reverses lifespan reduction by *spr-4;spr-3* mutations in *daf-2* mutant worms ($P = 1.1 \times 10^{-16}$ (*spr-4;spr-3;daf-2* +Ive versus −Ive)); *daf-2* −Ive, $n = 53$ worms; *daf-2* +Ive, $n = 55$; *spr-4;spr-3;daf-2* −Ive, $n = 95$; *spr-4;spr-3;daf-2* +Ive, $n = 69$; WT −Ive, $n = 64$. **d**, Multiple neurotransmitter and neuropeptide signalling systems contribute to the effects of *spr-4;spr-3* mutations on longevity. Change in lifespan of *spr-4;spr-3;daf-2* triple mutant and *daf-2* single mutant worms following neuronal RNAi of the indicated genes relative to empty vector control RNA. RNAi was targeted to neurons as described in Fig. 3b. $^{*}P < 0.05$, $^{**}P < 0.01$, Student's *t*-test. $n = 3$ independent experiments per group. Individual statistics are in Supplementary Table 22.

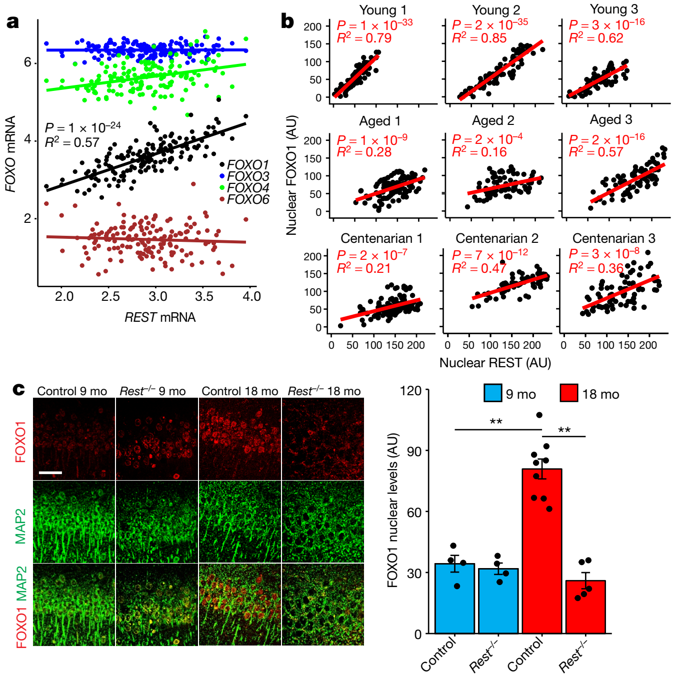


**Fig. 5 | REST regulates FOXO1 expression in the mammalian brain.** **a**, Linear regression analysis of *REST* and *FOXO1* mRNA expression in the prefrontal cortex of cognitively intact individuals (ROSMAP cohort age 71–101 years, $n = 150$) determined by RNA sequencing. *P* values derived by linear regression *t*-tests for the slope with Bonferroni correction for all expressed genes. **b**, Coordinate regulation of REST and FOXO1 in human prefrontal cortex. Nuclear REST and FOXO1 protein levels were determined by immunofluorescence microscopy in pyramidal neurons of the prefrontal cortex in individual young adults (20–38 years old), aged adults (70–80 years old), and centenarians (>100 years old). Each point represents a neuron that was double-labelled for REST and FOXO1. $n = 71$–114 neurons per individual. *P* values derived as in **a**. **c**, FOXO1 induction in the ageing mouse cortex is REST-dependent. Left, immunocytochemical labelling for FOXO1 and the neuronal marker MAP2 in cortical neurons of *Rest*$^{lx/lx}$ (control) and *Nestin-Cre;Rest*$^{lx/lx}$ (*Rest*$^{-/-}$) mice at 9 and 18 months (mo) of age. Scale bar, 40 μm. Right, quantification of FOXO1 nuclear levels. Mean ± s.e.m. Control 9 mo, $n = 4$ mice; control 18 mo, $n = 9$ mice; *Rest*$^{-/-}$ 9 mo, $n = 4$ mice; *Rest*$^{-/-}$ 18 mo, $n = 5$ mice. $^{**}P = 1.2 \times 10^{-5}$ control 9 mo versus control 18 mo; $^{**}P = 4.4 \times 10^{-7}$ control 18 mo versus *Rest*$^{-/-}$ 18 mo; ANOVA with post hoc Tukey test.

To identify the neural systems affected by SPR-4 and SPR-3, we used RNAi to target neurotransmitters and neuropeptides in *spr-4;spr-3;daf-2* triple mutants and *daf-2* single mutants. The most significant lifespan effects in the triple mutants were observed following RNAi directed against signalling through glutamatergic (*eat-4*), cholinergic (*cha-1/unc-17*) and monoaminergic (*cat-2*) neurons, and by RNAi against genes encoding axonal kinesin (*unc-104*) and the proprotein convertase (*egl-3*) (Fig. 4d, Extended Data Fig. 8i). These results suggest that SPR-3 and SPR-4 suppress excitatory neurotransmitter systems, as well as neuropeptide signalling, to extend lifespan in *daf-2* mutant worms.

Loss of function of *daf-2* extends lifespan through activation of DAF-16[22]. Following *daf-2* RNAi, *spr-4;spr-3* mutants showed reduced total and nuclear DAF-16 levels (Extended Data Fig. 9a). RNA-seq analysis of day10 *spr-4;spr-3;daf-2* triple mutant worms showed a reduction in the expression of DAF-16 target genes relative to *daf-2* single mutants (Extended Data Fig. 9b–d, Supplementary Tables 17, 18). Furthermore, inhibition of neural excitation by ivermectin restored levels of DAF-16 in *spr-4;spr-3* mutants following *daf-2* RNAi (Extended Data Fig. 9e). Thus, SPR-3 and SPR-4 suppress neural excitation (Fig. 4a, b), which leads to the activation of DAF-16.

## REST, FOXO1 and neural excitation

We next investigated whether the association between REST, neural excitation and human longevity (Fig. 2a–c) might be mediated by a mammalian forkhead transcription factor orthologous to *C. elegans* DAF-16. REST has been shown to regulate the expression of FOXO1 in SH-SY5Y neuroblastoma cells[15]. In the human brain, expression of *REST* mRNA was positively correlated with expression of *FOXO1* mRNA, but did not correlate with the expression of other FOXO family members (Fig. 5a, Extended Data Fig. 10a). Furthermore, REST and FOXO1 co-localized in neurons of the ageing human prefrontal cortex (Extended Data Fig. 10b), and nuclear levels of REST and FOXO1 were strongly positively correlated in all age groups (Fig. 5b).

To determine whether REST regulates FOXO1 expression in the brain, we examined *Rest* cKO and littermate control mice. FOXO1 localized predominantly to cortical neurons, and showed an age-related increase in 18-month-old relative to 9-month-old control mice (Fig. 5c). Age-dependent induction of nuclear FOXO1 was abolished in *Rest* cKO mice (Fig. 5c). Thus, regulation of forkhead transcription factors is a conserved feature of REST and its *C. elegans* orthologues.

To explore the role of neural excitation in regulation of FOXO1, we treated primary mouse cortical neuronal cultures with kynurenic acid (a broad-spectrum glutamate receptor antagonist) or with either 2-amino-5-phosphonovalerate (AP5) or 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[f]quinoxaline (NBQX) (antagonists for the NMDA (*N*-methyl-D-aspartate) and AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)/kainate subtypes of glutamate receptor, respectively). Kynurenic acid and NBQX significantly increased nuclear and total levels of FOXO1 (Extended Data Fig. 10c). Thus, FOXO1 is regulated by glutamatergic signalling in mammalian cortical neurons, paralleling the effect of neural excitation on DAF-16 in worms.

## Discussion

We have shown that extended longevity and cognitive preservation in humans is associated with coordinate downregulation of genes that mediate excitatory neurotransmission. In the model system *C. elegans*, an increase in the activity of excitatory ASH neurons is a normal aspect of ageing. Global inhibition of neural excitation, or inhibition of specific excitatory neuronal populations—particularly glutamatergic or cholinergic neurons—resulted in robust extension of lifespan. These findings are consistent with previous studies, which showed that the anticonvulsants ethosuximide and valproic acid can extend lifespan in *C. elegans*[23–25]. Moreover, we found that lifespan was dynamically regulated by the excitatory–inhibitory balance of neural circuits. Thus, an imbalance between neural excitation and inhibition might degrade neural function and contribute to the ageing process.

Our findings suggest that REST and the *C. elegans* orthologues SPR-3 and SPR-4 regulate ageing by acting as transcriptional repressors of synaptic genes and thereby reducing neural activity. Ageing conditional REST-deficient mice exhibit increased cortical neural activity and hyperexcitability. This is consistent with previous studies in neuronal cell culture, which suggested that REST maintains neural network homeostasis by buffering changes in neural excitation[26,27].

It is intriguing that REST and neural activity converge with insulin–IGF signalling to regulate the activity of forkhead transcription factors that play pivotal roles in lifespan regulation[22,28]. The activation of *daf-16* by REST orthologues in worms and *FOXO1* by REST in humans might be a mechanism for integration of neural activity with metabolism. We suggest that activation of REST and reduction of excitatory neural activity could be an approach to slowing ageing in humans.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1647-8.

Received: 28 November 2017; Accepted: 18 September 2019;

1. Satoh, A., Imai, S. I. & Guarente, L. The brain, sirtuins, and ageing. *Nat. Rev. Neurosci*. **18**, 362–374 (2017).
2. Bishop, N. A., Lu, T. & Yankner, B. A. Neural mechanisms of ageing and cognitive decline. *Nature* **464**, 529–535 (2010).
3. Ailion, M., Inoue, T., Weaver, C. I., Holdcraft, R. W. & Thomas, J. H. Neurosecretory control of aging in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **96**, 7394–7397 (1999).
4. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402**, 804–809 (1999).
5. Alcedo, J. & Kenyon, C. Regulation of *C. elegans* longevity by specific gustatory and olfactory neurons. *Neuron* **41**, 45–55 (2004).
6. Bishop, N. A. & Guarente, L. Two neurons mediate diet-restriction-induced longevity in *C. elegans*. *Nature* **447**, 545–549 (2007).
7. Wolkow, C. A., Kimura, K. D., Lee, M.-S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulinlike signaling in the nervous system. *Science* **290**, 147–150 (2000).
8. Lu, T. et al. Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
9. Mostafavi, S. et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci*. **21**, 811–819 (2018).
10. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci*. **19**, 1442–1453 (2016).
11. Gibbs, J. R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet*. **6**, e1000952 (2010).
12. Hilliard, M. A. et al. In vivo imaging of *C. elegans* ASH neurons: cellular response and adaptation to chemical repellents. *EMBO J*. **24**, 63–72 (2005).
13. Pokala, N., Liu, Q., Gordus, A. & Bargmann, C. I. Inducible and titratable silencing of *Caenorhabditis elegans* neurons in vivo with histamine-gated chloride channels. *Proc. Natl Acad. Sci. USA* **111**, 2770–2775 (2014).
14. Hamilton, B. et al. A systematic RNAi screen for longevity genes in *C. elegans*. *Genes Dev*. **19**, 1544–1555 (2005).
15. Lu, T. et al. REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507**, 448–454 (2014).
16. Hu, X. L. et al. Conditional deletion of NRSF in forebrain neurons accelerates epileptogenesis in the kindling model. *Cereb. Cortex* **21**, 2158–2165 (2011).
17. Brennan, G. P. et al. Dual and opposing roles of microRNA-124 in epilepsy are mediated through inflammatory and NRSF-dependent gene networks. *Cell Rep*. **14**, 2402–2412 (2016).
18. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
19. Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signallling in *C. elegans*. *Science* **298**, 830–834 (2002).
20. Yang, P. et al. A C-terminal truncated mutation of *spr-3* gene extends lifespan in *Caenorhabditis elegans*. *Acta Biochim. Biophys. Sin. (Shanghai)* **45**, 540–548 (2013).
21. Calixto, A., Chelur, D., Topalidou, I., Chen, X. & Chalfie, M. Enhanced neuronal RNAi in *C. elegans* using SID-1. *Nat. Methods* **7**, 554–559 (2010).
22. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. daf-16: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
23. Evason, K., Huang, C., Yamben, I., Covey, D. F. & Kornfeld, K. Anticonvulsant medications extend worm life-span. *Science* **307**, 258–262 (2005).
24. Evason, K., Collins, J. J., Huang, C., Hughes, S. & Kornfeld, K. Valproic acid extends *Caenorhabditis elegans* lifespan. *Aging Cell* **7**, 305–317 (2008).
25. Chen, X. et al. Ethosuximide ameliorates neurodegenerative disease phenotypes by modulating DAF-16/FOXO target gene expression. *Mol. Neurodegener*. **10**, 51 (2015).
26. Pozzi, D. et al. REST/NRSF-mediated intrinsic homeostasis protects neuronal networks from hyperexcitability. *EMBO J*. **32**, 2994–3007 (2013).
27. Pecoraro-Bisogni, F. et al. REST-dependent presynaptic homeostasis induced by chronic neuronal hyperactivity. *Mol. Neurobiol*. **55**, 4959–4972 (2018).
28. Li, Y. et al. Genetic association of FOXO1A and FOXO3A with longevity trait in Han Chinese populations. *Hum. Mol. Genet*. **18**, 4897–4904 (2009).

## METHODS

**Brain sample procurement and description.** Postmortem human brain material was procured in accordance with institutional guidelines and was approved by the Harvard Medical School Institutional Review Board. Tissue samples were procured from the Rush University Medical Center and the Brigham and Women's Hospital. Tissue samples (both paraffin-embedded and frozen) from Rush University Medical Center were derived from participants in the Religious Order Study (ROS) and Rush Memory and Ageing Project (MAP) (together referred to as ROSMAP) at the Rush Alzheimer's Disease Center; these are longitudinal, clinical–pathologic studies of ageing, cognitive decline and Alzheimer's disease[29,30]. Study participants agreed to comprehensive annual clinical and neuropsychological evaluation and to brain donation at death. Twenty-one cognitive function tests were used for the present study, including a summary score of all 17 tests used as a measure of global cognition, and separate measures of episodic, semantic, and working memory, perceptual speed, and visuospatial ability. The follow-up rate exceeds 95% and the autopsy rate exceeds 90%. All individuals who underwent autopsy cases were subject to a uniform structured neuropathologic evaluation of Alzheimer's disease, including assignment of Braak (measure of number and distribution of neurofibrillary tangles), CERAD (Aβ plaque pathology), and NIA-Reagan (composite measure of neurofibrillary tangles and amyloid plaques) scores (https://www.radc.rush.edu/docs/var/detail.htm?category=Pathology&subcategory=Alzheimer%27s%20disease&variable=ceradsc). Paraffin-embedded brain samples were also obtained from the Brigham and Women's Hospital. These samples included tissue from young adults without neurological abnormalities.

**Immunofluorescence analysis of human brain.** Immunofluorescence analysis of the prefrontal cortex (Brodmann areas 9, 10 and 47) was carried out using paraffin-embedded brain sections. Paraffin-embedded tissue sections were first deparaffinized in xylene, then rehydrated with decreasing concentrations of ethanol and placed in water. Sections then underwent antigen retrieval using the Diva decloaker (BioCare, USA). They were then washed and blocked with 2% BSA and 0.1% Triton X-100 in PBS for 1 h at room temperature. Primary antibodies were diluted in 2% BSA and 0.1% Triton in PBS. Following overnight incubation at 4 °C, sections were washed three times with PBS. Secondary antibodies, diluted in 2% BSA and 0.1% Triton in PBS were either biotin-coupled (1:200, Vector Labs, USA) or coupled to Alexa fluorophores (1:300, Invitrogen). Sections were incubated with 1% Sudan Black in 80% ethanol for 10 min at room temperature to suppress lipofuscin autofluorescence. Following washes in PBS, sections were mounted and imaged using confocal microscopy. The following antibodies were used for immunolabelling: (i) a rabbit polyclonal IgG that recognizes a region between residues 1050 and the C terminus (residue 1097) of REST (Bethyl laboratories, IHC-00141); (ii) a goat polyclonal IgG that recognizes the C-terminal region of FOXO1a (LSBio B415, discontinued, replaced with LSBio 1322). To quantify immunofluorescence, images that were randomly acquired in selected brain regions were analysed using the Metamorph image analysis system. Antigen-expressing areas within each neuron (such as the nucleus) were selected using the average signal intensity measured. Values were corrected by subtracting the average slide background intensity (measured outside cells). The investigator was blinded to sample origin or diagnosis.

To assess the relationship between the levels of nuclear REST protein and FOXO1, we performed confocal immunofluorescence with triple-labelling for REST, MAP2 and FOXO1. Multiple 40× pictures were acquired (at various locations) within the prefrontal cortex displaying pyramidal neurons using an Olympus Fluoview Confocal Microscope. For cases displaying a majority of pyramidal neurons with very high (or very low) nuclear REST levels, fields were also included that displayed lower (or higher, respectively) REST levels, to test for potential correlations between REST and FOXO1. Antigen-expressing areas within neuronal nuclei were selected using the Metamorph image analysis system and the average signal intensity was measured. Values were corrected by subtracting the average slide background intensity (measured outside cells). Between 70 and 115 pyramidal neurons were quantified for each case.

**Conditional *Rest* knockout mice.** Animal housing and experimental procedures were approved by the Institutional Animal Care and Use Committee of Harvard Medical School. Mice carrying floxed alleles of *Rest* flanking exon 2 were described previously[31,32]. These mice were crossed to *Nestin-Cre* transgenic mice (Jackson laboratory; strain 003771) to achieve conditional inactivation of *Rest* in the nervous system. The *Nes-Cre* transgene is in the C57BL/6J background, and the *Rest*$^{lx/lx}$ alleles were in a hybrid C57BL/6J and 129Sv/Ev background. The resulting *Nestin-Cre:Rest*$^{lx/lx}$ conditional knockout mice (hybrid C57BL/6J and 129Sv/Ev background), referred to as *Rest* cKO, were born at expected Mendelian ratios, were viable and fertile, and did not display any visible alterations. The control groups included *Rest*$^{lx/lx}$ and *Nes-Cre* mice. Mouse genotyping by PCR was performed using the following primers to amplify a region in the *Rest* gene flanking exon 2: re08 (5′-CATGCGAGTACTGCCATACCCAAC-3′), re09 (5′-GTGATGGGGCAGTCTTCTGGAGG-3′), and re11 (5′-GGCACACCTTTA ATCCTAGCTTC-3′); this allowed the identification of wild-type (220 bp), floxed (264 bp) or recombined (375 bp) *Rest* alleles. The experimental groups included both male and female mice in equal proportions. The *Rest* cKO and *Rest*$^{+/+}$ control groups were composed of littermate mice on the same genetic background. Mice were identified by eartag numbers, and were randomly selected for PET–CT and EEG experiments, as well as for histological processing (perfusion, brain dissection, and so on).

**Immunofluorescence analysis of mouse brain.** Mice were anaesthetized with isoflurane and carbon dioxide and then perfused with cold PBS buffer for 20 min. Brains were rapidly removed and placed in 4% PFA overnight at 4 °C. They were then processed for paraffin embedding, according to standard procedures. The investigator was blind to the genotype. To assess FOXO1 nuclear expression in cortical neurons, coronal brain sections from wild-type and *Rest* cKO mice (aged 9 or 18 months) were immunolabelled with FOXO1 and MAP2 antibodies. Cortical neurons (MAP2$^{+}$) were identified and the mean FOXO1 fluorescence intensity in each nucleus was measured using Metamorph software. Between 50 and 100 neuronal nuclei were assessed for FOXO1 expression, and the mean FOXO1 nuclear expression was derived for each animal. To confirm REST deficiency (Extended Data Fig. 5e), *Rest* cKO mouse cortical sections were labelled with an anti-REST antibody provided by the Hsieh laboratory (Supplementary Table 20).

**PET–CT of *Rest* cKO mice.** Mice were anaesthetized with 3% isoflurane (Baxter Medical) and medical grade oxygen at a rate of 1 l/min. A CT scout scan was done first, followed by a CT scan and a dynamic PET scan. Each mouse received the same dose per gram of body weight (1.75 Ci/g) of $^{18}$fluorodeoxyglucose (FDG) tracer solution by tail vein injection, followed by a 0.1 ml saline flush. Dynamic PET imaging for each mouse was immediately performed for 1 h (in vivo PET) using a small animal PET/CT scanner (eXplore Vista; GE Healthcare). The spatial resolution of the PET scanner was 1.6 mm at the centre of the field of view. The data were acquired in 3D mode at the energy window of 250–700 keV, which yields 4% count sensitivity. For each time point, five or six 3D volumes, spanning cortical and subcortical regions, were selected in the centre of the brain, with volumes of 0.2 cm$^3$ each, and used for quantification using the eXplore Vista software. The averages of these regions were used as the SUV for the animal.

**Cell culture.** Primary cortical neuronal cultures, derived from E16.5 wild-type C57BL/6J fetuses, were plated in 10% serum-containing neuronal culture medium (neurobasal medium containing B27 supplements, penicillin, streptomycin, and GlutaMax) on either coverslips or culture dishes that were pre-coated with poly-L-ornithine (Sigma p4957). The medium was changed 4 h after initial plating to serum-free neuronal culture medium, and then a half-medium change was performed every three days.

**Electroencephalography of *Rest* cKO mice.** *Electroencephalogram (EEG) telemetry unit implantation.* Mice were implanted with wireless telemetry units (PhysioTel ETA-F10; Data Sciences International, DSI, St. Paul, MN) under appropriate sterile techniques per laboratory protocol as previously described[33–36]. Mice were anaesthetized by intraperitoneal (i.p.) injection of 100 mg/kg ketamine (Putney Vet, Portland, ME) and 10 mg/kg xylazine (Lloyd Inc., Shenandoah, IA). The transmitter was placed intraperitoneally, and electrodes were threaded subcutaneously to the cranium. After skull exposure, haemostasis, and identification of the cranial sutures, bregma and lambda, two 1-mm diameter burr holes were drilled over the right olfactory bulb (reference) and left occipital cortex (active). The epidural electrodes of the telemetry units, connected to the leads of the transmitter, were placed into the burr holes and secured using stainless steel skull screws. Once in place, the skull screws were covered with dental cement (Dentsply International Inc., Milford, DE). Mice were subcutaneously injected 0 and 24 h post-operatively with 5 mg/kg meloxicam (Norbrook Laboratories, Newry, Northern Ireland) for analgesia. After 1 week of recovery, mice were individually housed in their home cages in a 12 h light–12 h dark, temperature, and humidity controlled chamber with ad libitum access to food and water in preparation for recording.

*Video EEG recording, seizure induction and analysis.* One-channel video–EEG was recorded differentially between the reference (right olfactory bulb) and active (left occipital lobe) electrodes after 24 h of acclimation in recording chambers. EEG (1,000 Hz), core-body temperature and locomotor activity (200 Hz) signals were continuously sampled over a period of 48 h along with time-registered videos. At the end of baseline EEG acquisition, all mice were provoked with a convulsive dose (40 mg/kg; i.p.) of the GABAA receptor antagonist PTZ (Sigma-Aldrich, Co., St. Louis, MO) to measure seizure susceptibility and evaluate seizure thresholds[33,36]. Mice were continuously monitored for clinical and electrographic epileptiform activity during both periods and this was post hoc verified by blinded review of the video EEG. Following PTZ administration, latency to generalized tonic-clonic seizures (GTCs), number of seizures, and total duration of GTCs were recorded for each mouse. Mice without seizures were assigned a time of 20 min at the end of the PTZ challenge observation period. One *Rest* cKO mouse was excluded from the analysis based on a febrile temperature of 40.68 °C, compared with an average temperature of 34.56 ± 1.48 °C for all other mice.

**Immunocytochemical analysis of cultured cells.** Embryonic mouse cortical neuronal cultures as described above were plated on poly-L-ornithine-coated coverslips. Stocks of NBQX, APV(Tocris Bioscience Cat. No. 0190, 0106, respectively), and kynurenic acid sodium salt (Abcam, ab146693) were added to neurobasal medium and used in a half-medium change for a final concentration of 5 μM kynurenic acid, 50 μM APV or 2 μM NBQX at day 10 of culture. After 24 h, the culture medium was aspirated and cells were fixed by incubation with 4% (v/v) paraformaldehyde in PBS for 20 min at room temperature, and permeabilized with 0.2% Triton X-100 in PBS for 15 min at room temperature. After a wash in PBS, cells were blocked with 4% BSA in PBS overnight at 4 °C. Primary antibodies were diluted to the appropriate concentration in 4% BSA, and incubated with the cells overnight at 4 °C. Cells were then washed three times in PBST (PBS with 0.05% Triton) for 10 min each, before we added fluorophore-conjugated secondary antibodies for 2 h at room temperature. Fluorophore-labelled cells were then washed in PBST (3 × 10 min) and mounted using Prolong Gold mounting medium with DAPI and anti-fade reagent (Invitrogen). The primary antibodies used in Fig. 5d are rabbit anti-FOXO1 (Cell Signaling 2880) and chicken anti-MAP2 (Sigma-Millipore, AB5392). Nuclei were identified by DAPI labelling, and the cellular distribution of FOXO1 labelling was quantified using Metamorph software. Cell identification parameters were optimized for scoring neuronal cultures, and then the same image analysis macro was applied to all images to generate average FOXO1 cellular intensities in MAP2-positive neurons.

***C. elegans* strains.** The N2 Bristol strain was used as the wild-type background for these studies. *C. elegans* strains were cultured at 20 °C under standard conditions as described[37]. The following mutations and chromosome rearrangements were used: LGI: *spr-4(by105)*[38], *spr-4(tm465)*[15], *daf-16(mu86)*[39], *hT2[bli-4(e937) qIs48] (I;III), unc-13(e51)*[37]; LGIII:*daf-2(e1370)*[39]*eat-4(nj2)*[40]; LGV: *spr-1(ok2144)*[41], *sid-1(pk3321)*[42], *egl-3(gk238)*[43]and LGX *spr-3(ok2525)*[15]; EG7215(oxTi334 [*eft-3p*::*TdTomato*::*h2b*::*unc-54 3′UTR* + *Cbr-unc-119*(+)]). All strains, except for *spr-4(tm465)*, were provided by the Caenorhabditis Genetics Center (CGC), which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440). Worms that had not been outcrossed previously were outcrossed six times before use. See Supplementary Table 19 for a full listing of strains.

The *spr-5* mutants exhibited transgenerational effects, including increased defects in DNA double-strand break repair and germline apoptosis[44] as well as lifespan effects[45]. Although we have not fully evaluated the transgenerational phenotypes of *spr-3* and *spr-4* alleles, as a precaution, all lifespan assays were performed with worms no more than five generations removed from a heterozygote ancestor. To facilitate this, *spr-4* alleles were carried over the hT2 balancer described above, while *spr-3*(*ok2525*) was carried over the MosTIC[46] insertion *oxTi335*, which carries a *eft-3::tdTomato::h2b* fluorescent fusion construct integrated into the X chromosome at 4,348,071 bp. As *spr-3* is located on the X chromosome at 4,514,000 bp, this effectively marked wild-type *spr-3*, allowing us to maintain heterozygous stocks of our double and single mutants.

The *daf-16::gfp* fusion line TJ356, bearing transgene *zIs356* [*daf-16p::daf-16a/b*::GFP + *rol-6*], was mildly anaesthetized with 0.01% tetramizole and outcrossed to laboratory wild-type males twice, and then mated to *spr-4(by105);spr-3(ok2525)* double mutants. Double mutants bearing *zIs356* were recovered and propagated.

***C. elegans* RNAi.** Feeding RNAi experiments were performed at 20 °C as described[47]. A feeding clone containing a full length *daf-2* cDNA was provided by K. Blackwell. A *daf-16* RNAi vector was provided by E. Greer. Control RNAi was performed by feeding HT115 bacteria carrying the empty pL440 vector. A list of all RNAi clones used is available in Supplementary Table 20.

Enhanced neuronal RNAi was achieved as described[21], by using *sid-1(pk3321)* mutants in which *sid-1* (a dsRNA transporter required for RNAi) is re-expressed solely in neurons via a *punc-119*:*sid-1* transgene. The animals are therefore systemically deficient in RNAi in all cell types except neurons, where RNAi is enhanced by heightened expression of *sid-1*.

**Microinjection and transgenic strains.** For transgenic *C. elegans* experiments, lines were generated by microinjecting the relevant constructs into the gonads of the indicated worm strains. *Prab-3*::mCherry (pGH8, Addgene: 19359) or *Pmyo-2*::mCherry (pCFJ90, Addgene 19327) were used as coinjection markers. Three independent lines that demonstrated reliable transmission of the marker were propagated as described above for each experiment. Where possible, non-array segregating worms were also maintained as controls. For the dCas9::VP64 experiments, a *peft-3::dCas9VP64::tbb-2* UTR construct was cloned using the Wormgate Gateway recombination system into the pCFJ151 MosSCI destination vector. The EG6699 strain, bearing the *ttTi5605*, MosSCI integration site on chromosome II (at 8.42 MB) was raised on HT115 and microinjected as described[48]. Integration was confirmed by genotyping, as well as the absence of unrescued *Unc*$^-$ progeny. These worms were then microinjected with a cocktail of four sgRNAs (5 ng/μl each) generated by nested overlap PCR (see Supplementary Table 21 for sequences of sgRNAs), along with a pGH8 *myo-2::mcherry* (10 ng/μl) as a marker. Three transgenic lines from independent injections were selected. For simplicity of presentation, these are presented as merged data for lifespans and qPCR validation. For the GCaMP experiment, two of the lines were selected for mating and analysis.

The SPR-4::GFP reporter line was generated by microparticle bombardment using a fusion construct and bombardment protocol as described[49].

**Transcript analysis in *C. elegans*.** To analyse RNA from transgenic lines, 100 worms (24 h post-L4) of each strain (including controls) were placed into 1.5 ml M9 buffer. For RNA-seq experiments, 450 worms per genotype were grown on FUDR containing plates (starting at L4) and collected in M9, either on day 2 or day 10 of adulthood. Worms were washed once in M9, pelleted by centrifugation, resuspended in 200 μl Trizol, vortexed for 2 min and flash frozen in liquid nitrogen. Worms were then freeze-cracked by thawing in a 37 °C water bath and re-freezing in liquid nitrogen. This was repeated two more times. After the final thaw, 100 μl Trizol was added and the tubes were maintained at room temperature for 5 min. RNA was then extracted with 140 μl chloroform, precipitated with an equal volume of 70% ethanol and transferred to an RNeasy spin column (Qiagen) and purified. Quantitative RT–PCR was performed directly from isolated RNA, using 1 ng of RNA and the Qiagen One Step qPCR mix. All reactions were performed in triplicate. The ddCt method was used to analyse qRT–PCR data, and dCt values were used for statistical analysis.

***C. elegans* motility assay.** Worms were treated for 24 h with the indicated drugs, and then 150 worms were transferred to 1.5 ml liquid nematode growth medium (NGM; 1 mM MgCl$_2$, 1 mM CaCl$_2$, 200 mM KH$_2$PO$_4$, 50 mM NaCl) and washed once to remove bacterial clumps. Worms were transferred in 100-μl volumes to a 96 u-shaped-well plate, and assayed in the Nemametrix wMicrotracker (https://nemametrix.com/product-category/phenotyping-products/wmicrotracker/) according to the manufacturer's instructions.

***C. elegans* lifespan determination and stress treatments.** Lifespan and ageing experiments were performed at 20 °C and on fluorodeoxyuridine (FUdR) unless otherwise noted. For each genotype, 20–35 day 1 worms were transferred to NGM plates containing 100 μg/ml FUdR. Worms were scored for viability every day or every other day, and transferred to fresh plates between day 10 and day 14. For *daf-2* mutant worms, worms were moved again at day 20–24. For histamine and ivermectin experiments, where the same plates were used for the entirety of the lifespan, a wetted towel was placed in the container with the plates to mitigate evaporative loss of water from the plates. For lifespan experiments in which untreated worms were transferred to the treatment group (for example, day 8 histamine treatment), control worms with or without treatment were also transferred to fresh plates of the same type to control for any lifespan effects resulting from plate transfer. The presence or absence of FUdR is indicated for each experiment in Supplementary Table 22.

For all lifespans and stress resistance experiments, worms that burst (interior leaking out through the vulva) or bagged (interior hatching of progeny) were discarded from the plate and not used in lifespan analysis, along with animals that crawled off the plate during the course of the assay. Plates with mould or other contamination were discarded.

**Quantification of DAF-16::GFP fluorescence.** For quantification of GFP fluorescence in fixed animals, 25–30 worms were transferred to an Eppendorf tube containing M9 buffer, washed once in M9 and pelleted. The supernatant was removed and the pellet was frozen in liquid nitrogen. The day before imaging, worm pellets were thawed in PBS with 4% paraformaldehyde (ThermoFisher) and fixed for 30 min at room temperature while rocking. The PFA was removed by washing twice with PBS, with 0.025% Triton (PBST, to prevent sticking), and the worms were stained with DAPI for 10 min at room temperature. The DAPI was removed with a final wash in PBST, and the pellet was resuspended in 15 μl ProLong Gold mounting medium (Life Technologies). The worms were transferred in 12.5 μl mounting medium to a slide, and gently placed under a coverslip, which was then sealed with clear nail polish (Electron Microscopy Sciences).

Slides were imaged on an Olympus Fluoview 1000 confocal microscope, using manufacturer settings for GFP and DAPI (488 nm and 405 nm laser, respectively) within 2 weeks of mounting, and all slides compared with each other were imaged on the same day, in the same imaging session. Worms were imaged via a 40× objective with a 2× digital zoom. For every worm, *Z*-stack image series were taken at 2-μm step sizes (roughly 13–15 images per series). To quantify nuclear GFP levels, these *Z*-stacks were opened in ImageJ (NIH) and flattened using a maximum intensity *Z*-stack projection. For pharyngeal nuclei, three regions of interest (ROI) were selected on the basis of DAPI staining of nuclei, on either side of the pharynx and behind it, encompassing the bulk of the nuclei in that area. These were then used as the ROIs within which to measure GFP fluorescence, using the ImageJ intensity measuring tool. Average scores from the ROIs were computed for each worm. For Extended Data Fig. 6b (middle), Extended Data Fig. 7d (right) and Extended Data Fig. 9a (right), all nuclei located in the middle three *z* planes of the worm were individually selected using the same ROI and intensity measuring tools. These nuclei were then used to calculate an average score per worm.

**GCaMP imaging in *C. elegans*.** *Data acquisition.* GCaMP imaging was performed in lines bearing *kyIs602* [sra-6::GCaMP3.0, 75 ng/µl + unc-122::GFP, 10 ng/µl], which expresses predominately in the ASH neurons. Worms were removed from plates and mounted on 7.5% agarose pads in liquid NGM (1 mM $MgCl_2$, 1 mM $CaCl_2$, 200 mM $KH_2PO_4$, 50 mM NaCl), mixed 1:1 with 0.05-µm polystyrene beads (Polysciences, Inc., cat: 08691) as described (http://wbg.wormbook.org/2009/12/01/agarose-immobilization-of-c-elegans/). A coverslip was very gently applied and worms were imaged for a maximum of 30 min after mounting. Videos lasting 144 s (240 frames, ~0.6 frames per second) were recorded on the confocal microscope described above (FV1000), with the confocal aperture widened to 250 µm to allow lower excitation intensity and mitigate *z*-drift. Worms were imaged with factory GFP settings and the 488-nm laser set to 2%. Videos were opened and played in Image J, and excitation events were scored manually for Fig. 4a. Scoring was performed blinded to genotype.

*Analysis of GCaMP intensity changes.* For analyses of $\Delta F/F_0$ maximum intensity, videos were analysed using a custom MATLAB script, which automatically registered and segmented the ASH neuron and recorded GCaMP fluorescence intensity for each frame. Videos of tracked neurons were manually reviewed and neurons that were poorly tracked were excluded from analysis. To normalize the fluorescence intensity per worm we used $\Delta F/F_0$, calculated as $(F(t) - F_0)/F_0$ where $F(t)$ is the fluorescence intensity at time $t$ and $F_0$ represents an estimate of the baseline fluorescence, calculated as the 0.2 quantile fluorescence per worm. $\Delta F/F_0$ maximum intensity was the maximum $\Delta F/F_0$ over the entire recording (240 frames). For analysis of nemadipine-treated worms and for worms expressing the HisCl channel (and their controls), ROIs were hand-drawn because the tracking program failed, owing either to low signal intensity (nemadipine treated worms) or the inability to distinguish a confounding GFP signal (HISCL::GFP worms). Some worms were used in multiple GCaMP analyses.

***C. elegans* cloning and genotyping.** The SPR-4::GFP fusion fosmid used in bombardment (Extended Data Fig. 6a, b; Clone: 3167840880351681 C09) was provided by the Transgenome consortium[49]. All primers used for genotyping can be found in Supplementary Table 21.

**Human brain gene expression cohorts.** *ROSMAP cohort.* Brain samples were from the dorsolateral prefrontal cortex of 638 individuals spanning the range of cognitive function from cognitively normal to MCI to AD from the ROS and MAP cohorts (https://www.synapse.org/#!Synapse:syn3219045).

*CommonMind Consortium cohort.* Brain samples were from the dorsolateral prefrontal cortex of 602 individuals with schizophrenia or bipolar disease and control individuals with no neuropsychiatric disorders (https://www.synapse.org/#!Synapse:syn2759792).

**Gibbs cohort.** Brain samples were from the frontal cortex of 146 neurologically normal individuals[11].

**RNA library preparation and sequencing.** *Human RNA-seq.* Details of RNA-seq for the ROSMAP cohort can be found at https://www.synapse.org/#!Synapse:syn3388564; details of RNA-seq for the CMC can be found at https://www.synapse.org/#!Synapse:syn3157743.

C. elegans *day 2.* Libraries were prepared using Illumina TruSeq Stranded mRNA sample preparation kits from 500 ng of purified total RNA according to the manufacturer's protocol. The finished dsDNA libraries were quantified using Qubit fluorometer, Agilent TapeStation 2200, and RT–qPCR using the Kapa Biosystems library quantification kit according to the manufacturer's protocols. Uniquely indexed libraries were pooled in equimolar ratios and sequenced on an Illumina NextSeq 500 with paired-end 75-bp reads by the Dana-Farber Cancer Institute Molecular Biology Core Facilities.

C. elegans *day 10.* cDNA was synthesized using Clontech SmartSeq v4 reagents from 2 ng RNA, owing to low yields from aged worms. Full-length cDNA was fragmented to a mean size of 150 bp with a Covaris M220 ultrasonicator and Illumina libraries were prepared from 2 ng sheared cDNA using Rubicon Genomics Thruplex DNaseq reagents according to the manufacturer's protocol. The finished dsDNA libraries were quantified using Qubit fluorometer, Agilent TapeStation 2200, and RT–qPCR using the Kapa Biosystems library quantification kit. Uniquely indexed libraries were pooled in equimolar ratios and sequenced on Illumina NextSeq500 run with paired-end 75-bp reads at the Dana-Farber Cancer Institute Molecular Biology Core Facilities. One sequenced sample consisted of primarily ribosomal RNA and was discarded. This sample was reprocessed along with two other previously sequenced samples (as quality controls) with a new library preparation and sequencing. On a principal components plot these three samples showed no clear batch effect compared to the other samples so the resequencing of the sample that was problematic in the original batch was used. The two additional resequenced samples used for quality control were not used in the analysis.

**Human sample selection and outlier analysis.** *ROSMAP cohort.* Samples were excluded if missing the cognitive diagnosis (cogdx) or post-mortem interval (PMI), if the expression of sex-specific genes did not agree with reported sex, or if RNA integrity (RIN) was less than 6.5. Of the 638 samples with RNA sequencing we removed three because of missing information, 180 because of low RIN, one owing to sex mismatch, and the one remaining sample from sequencing batch 7, the other samples of which had a strong batch effect on principal components plots and were removed because they had low RIN. Next we performed outlier analysis (see below), and removed ten outliers (three no cognitive impairment (NCI), one mild cognitive impairment (MCI), five Alzheimer's disease (AD) dementia, one other dementia). After quality control, 443 individuals remained for analysis (150 NCI, 118 MCI, 164 AD dementia, 11 other dementia).

*CommonMind Consortium cohort*[10]. We excluded the ten outlier samples identified in the original publication. No other samples were removed. After quality control, of the 602 samples with RNA sequencing, there were 592 individuals for analysis (279 control, 258 schizophrenia, 47 bipolar, 8 affective or mood disorder).

*Gibbs cohort.* No samples were removed because of missing information or incorrect sex. Next we performed outlier analysis, and removed two outliers. After quality control, there were 144 individuals for analysis.

*Outlier analysis.* Outlier analysis was performed for each cohort in an iterative manner using only genes expressed in the 'youngold' and 'oldestold' groups at each iteration (see sections Gene expression sample groups for human cohorts and *C. elegans* and Gene expression normalization and covariate adjustment). For the ROSMAP and CMC cohorts we used edgeR TMM-normalized log(counts per million (CPM)) expression. For the Gibbs cohort we used $\log_2$ expression. At each iteration, the sample deemed to be the strongest outlier, if any, was removed and the process was iterated until no outliers were detected. Outlier detection was performed using principal components analysis, agglomerative hierarchical clustering using the Euclidean distance metric and average linkage criterion, and pairwise correlation of gene expression per sample. Samples were classified as outliers if they were isolated on a plot of the first and second principal components or in the hierarchical clustering tree, and had relatively low correlation with other samples determined as its mean pairwise correlation with other samples more than 3 s.d. below the mean.

**RNA sequencing read alignment and quantification of gene expression.** *ROSMAP cohort.* BAM files containing reads aligned to hg19 were obtained from synapse (https://www.synapse.org/#!Synapse:syn3388564). Reads were obtained from these BAM files using Picard v1.138 (http://broadinstitute.github.io/picard/) SamToFastq. Reads were aligned to GRCh38 with Ensembl GRCh38.86 gene models using STAR version 2.5.2b[50] with options–outFilterMismatchNoverLmax 0.04 –outFilterMismatchNmax 999 –alignSJDBoverhangMin 1 –alignSJoverhangMin 8 –outFilterMultimapNmax 20 –outFilterType BySJout –alignIntronMin 20 –alignIntronMax 1000000 –alignMatesGapMax 1000000. The expression of genes was quantified as gene counts using FeatureCounts v1.5.1[51] with options -C -p -B -s 2 -t exon -g gene_id.

*CommonMind Consortium cohort.* Gene counts were obtained from synapse (https://www.synapse.org/#!Synapse:syn3346749).

C. elegans *day 2 and day 10.* Quality control of sequencing reads was performed with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the *C. elegans* WBcel235 genome with Ensembl WBcel235.86 gene models using STAR version 2.5.2b with options –outFilterMismatchNoverLmax 0.04 –outFilterMismatchNmax 999 –alignSJDBoverhangMin 1 –alignSJoverhangMin 8 –outFilterMultimapNmax 20 –outFilterType BySJout –alignIntronMin 15 –alignIntronMax 1000000 –alignMatesGapMax 1000000. The expression of genes was quantified as gene counts using FeatureCounts v1.5.1 with options -C -p -B -s 2 -t exon -g gene_id for day 2 and -C -p -B -s 0 -t exon -g gene_id for day 10. Data are available in the Gene Expression Omnibus (GEO) under accession number GSE123146.

**Gene expression sample groups for human cohorts and *C. elegans*.** *ROSMAP cohort.* ROSMAP was our discovery cohort. On the basis of hierarchical clustering of age-associated genes (see section Hierarchical clustering to select age group cutoffs), we partitioned individuals with no cognitive impairment into the age groups: youngold (28 individuals, ages $\geq$70 and $\leq$80 years), middleold (33 individuals, ages >80 and <85 years), and oldestold (89 individuals, ages 85–101 years). Individuals with cognitive impairment were grouped into a separate cognitive impairment group (293 individuals).

*CommonMind Consortium cohort.* The CMC cohort was a replication cohort for the ROSMAP results. We partitioned control individuals into four age groups: young (105 individuals, ages $\geq$17 and $\leq$60 years), youngold (96 individuals, ages $\geq$60 and $\leq$80 years), middleold (19 individuals, ages >80 and <85 years) and oldestold (59 individuals, ages $\geq$85 and 90+ years; in this cohort, age of death greater than 90 years was censored to 90+). Individuals with neuropsychiatric illnesses were grouped into a separate group (313 individuals).

*Gibbs cohort.* The Gibbs cohort was a replication cohort for the ROSMAP results. We partitioned cognitively normal individuals into four age groups: young (104 individuals, ages $\geq$15 and <55 years), youngold (18 individuals, ages $\geq$55 and $\leq$80 years), middleold (3 individuals, ages >80 and <85 years), and oldestold (19 individuals, ages 85–101 years).

Throughout the paper, 'individuals with extended longevity' refers to individuals in the oldestold groups defined above.

C. elegans *day 2 and day 10.* There were four genotypes of worm: *daf-2*, *daf-2;spr-4;spr-3*, N2, and *spr-4;spr-3*, and each genotype had three biological replicates. Each genotype was considered as a separate group for analysis.

**Gene expression normalization and covariate adjustment.** *ROSMAP cohort.* Gene counts were input to edgeR[52]. Genes were deemed to be expressed if $\geq$10 individuals in a combined youngold and oldestold group had $\geq$1 CPM. Genes not satisfying these criteria were removed, keeping the original library sizes. These were expressed genes used during each iteration of outlier analysis. After outlier analysis, this filtering retained 18,511 expressed genes out of 58,051 annotated genes for analyses involving the ROSMAP cohort. Counts were then normalized using the TMM method in edgeR. Finally, log(CPM) values were calculated for analyses other than differential expression.

To adjust gene expression for covariates we fit the linear regression model for each gene separately using lm() in R: gene expression ~ group + covariates where gene expression is log(CPM), and using the group and covariates for ROSMAP. For ROSMAP, group was a factor with four levels: youngold, middleold, oldestold, and cognitivedecline, with cognitivedecline as the reference level. The covariates were sex (factor, two levels), RIN (continuous), RIN^2 (continuous), PMI (continuous), and sequencing batch (factor, eight levels). The final normalized and adjusted gene expression values were derived from adding the regression residuals to the estimated effect of the group level to preserve the effect of the group on expression. These normalized and adjusted gene expression values were used to perform gene–gene regression analysis and gene–gene group regression analysis, and to visualize gene expression.

*CommonMind Consortium cohort.* Gene counts were input to edgeR. Genes were deemed expressed if $\geq$10 individuals in a combined youngold and oldestold group had $\geq$1 CPM. Genes that did not satisfy these criteria were removed, maintaining the original library sizes. These were expressed genes used during each iteration of outlier analysis. After outlier analysis, this filtering retained 19,453 expressed genes out of 56,632 annotated genes for analyses involving the CMC cohort. Counts were then normalized using the TMM method in edgeR. Finally, log(CPM) values were calculated for analyses other than differential expression. To adjust gene expression for covariates, we fit the linear regression model for each gene separately using lm() in R: gene expression ~ group + covariates where gene expression is log(CPM), and using the group and covariates for CMC. For CMC, group was a factor with five levels: young, youngold, middleold, oldestold, and neuropsychiatric illness, with neuropsychiatric illness as the reference level. The covariates used were selected from those used in the original publication: sex (factor, two levels), RIN (continuous), RIN^2 (continuous), PMI (continuous), clustered batch (factor, nine levels), and institute (factor, three levels). The final normalized and adjusted gene expression values were derived by adding the regression residuals to the estimated effect of the group level to preserve the effect of the group on expression. These normalized and adjusted gene expression values were used to perform gene–gene regression analysis and gene–gene group regression analysis, and to visualize gene expression.

*Gibbs cohort.* Gene expression data measured on the Illumina humanRef-8 v2.0 expression beadchip platform were downloaded from NCBI GEO GSE15745. Raw intensity values for each probe were transformed using the rank invariant normalization method by the authors[11], and then $\log_2$ transformed for analysis. Probes were deemed expressed if $\geq$10 individuals in a combined youngold and oldestold group had detection $P < 0.01$. Unmapped probes according to updated GEO Platform (GPL) annotation were removed. These were expressed probes used during each iteration of outlier analysis. After outlier analysis, this filtering retained 13,239 expressed probes out of 22,184 probes for analyses involving the Gibbs cohort.

To adjust gene expression for covariates we fit the linear regression model for each gene separately using lm() in R: probe expression ~ group + covariates where probe expression is $\log_2$ expression, and using the group and covariates for Gibbs. For Gibbs, group was a factor with four levels: young, youngold, middleold, and oldestold, with young as the reference level. The covariates were sex (factor, two levels), PMI (continuous), and prep hybridization batch (factor, seven levels). The final normalized and adjusted probe expression values were derived by adding the regression residuals to the estimated effect of the group level to preserve the effect of the group on expression. These normalized and adjusted probe expression values were used to visualize probe expression.

C. elegans *day 2 and day 10.* Gene counts were input to edgeR. Genes were deemed expressed if $\geq$3 samples had $\geq$1 CPM. Genes not satisfying these criteria were removed, keeping the original library sizes. This filtering retained 12,981 expressed genes out of 46,739 annotated genes for analyses involving *C. elegans* day 2 and 15,154 expressed genes for analyses involving *C. elegans* day 10. Counts were then normalized using the TMM method in edgeR. Finally, log(CPM) values were calculated for analyses other than differential expression. For day 2, covariate adjustment was not necessary because there were no other covariates beyond the group variable. For day 10 we did not include a batch covariate because there was no observed batch effect on a principal components plot, and results were similar if a batch covariate was included and the two dropped technical repeat sequencings were used instead of their original sequenced reads (data not shown). Thus, covariate adjustment was not necessary because there were no other covariates beyond the group variable. These normalized gene expression values were used to visualize gene expression.

To visualize differentially expressed genes in heat maps, normalized gene expression values were transformed to a *z*-score per gene and thresholded to $[-3, 3]$.

**Differential expression analysis.** *ROSMAP cohort.* Differential expression analysis between the oldestold and youngold groups with covariate adjustment using the covariates listed above for ROSMAP was performed for expressed genes using edgeR (estimateDisp, glmFit, and glmLRT with default arguments) in R. All 443 samples were included to increase statistical power during covariate modelling. Genes were considered differentially expressed if the Benjamini and Hochberg false discovery rate (FDR) was $\leq$0.1, which was calculated using the R function p.adjust with argument method = 'fdr'[53].

*CommonMind Consortium cohort.* Differential expression analysis between the oldestold and youngold groups with covariate adjustment using the covariates listed above for the CMC was performed for expressed genes using edgeR (estimateDisp, glmFit, and glmLRT with default arguments) in R. All 592 samples were included to increase statistical power during covariate modelling. Genes were considered differentially expressed if FDR $\leq$ 0.05 and the absolute value of the fold change was $\geq$1.2.

*Gibbs cohort.* Differential expression analysis between the oldestold and youngold groups with covariate adjustment using the covariates listed above for Gibbs was performed for expressed probes using linear regression models in limma (lmFit and eBayes with default arguments) in R. All 144 samples were included to increase statistical power during covariate modelling. Probes were considered differentially expressed if FDR $\leq$ 0.05.

C. elegans *day 2 and day 10.* Differential expression analysis between all pairs of groups such as *spr-4;3;daf-2* versus *daf-2* was performed for expressed genes using edgeR (estimateDisp, glmFit, and glmLRT with default arguments) in R. All 12 samples were included during edgeR analysis. Genes were considered differentially expressed if FDR $\leq$ 0.05.

**Gene sets and gene set enrichment analysis.** Gene annotation and gene sets used for functional gene classification into biological process (BP), molecular function (MF), and cellular component (CC) were from the GRCh38.p7 database downloaded from Ensembl Biomart (https://www.ensembl.org/biomart/martview/) on November 2, 2016.

Synaptic transmission genes were genes that had direct or indirect annotation for the GO biological process 'Chemical synaptic transmission' (GO: 0007268).

To identify REST target genes, the REST RE1 motif position-specific weight matrix MA0138.2 was obtained from JASPAR[54]. FIMO version 4.10.1[55] was used with the *Homo sapiens* genome sequence GRCh38 to predict REST binding sites. A gene was defined to be a REST target if it had an RE1 motif with motif $P < 1 \times 10^{-7}$ that was ±10 kb from the transcription start site of any transcript of the gene in the Ensembl GRCh38.86 gene models. This procedure identified 2,632 REST target genes before filtering for expressed genes in each cohort during gene set enrichment analysis.

ENCODE ChIP–seq transcription factor target gene sets were from the ENCODE_TF_ChIP-seq_2015 database[56] downloaded from http://amp.pharm.mssm.edu/Enrichr/ on 8 March 2018.

Cell type analysis gene sets were derived from a transcriptome database of the major cell classes of the mouse cerebral cortex[57] using data from https://web.stanford.edu/group/barres_lab/brain_rnaseq.html. To select cell marker genes, first we calculated the fold expression of each gene in each cell type by dividing the fragments per kilobase of transcript per million mapped reads (FPKM) expression in that cell type by the mean of the FPKM expression in the other six cell types. For each cell type we selected as cell marker genes those that had at least tenfold higher expression in that cell type and that had one-to-one human–mouse homologues. Homologues were downloaded from http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt on November 16, 2016.

DAF-16 class I (upregulated) genes were derived by selecting genes with FDR $< 0.01$ and class == 'up' from Supplementary Table 1 of Tepper et al.[58].

For gene set enrichment analysis of differentially expressed genes, we retained only genes in each gene set that were expressed in that differential expression analysis. Gene set enrichment analysis was performed separately for upregulated genes and non-upregulated genes, downregulated genes and non-downregulated genes, and differentially regulated genes and non-differentially regulated genes. For Gene Ontology gene sets, gene set enrichment was determined using the topGO R package[59] using the classic algorithm and Fisher's statistic. Gene set enrichment for REST target genes, ENCODE ChIP–seq transcription factor target gene sets, and cell type analysis gene sets was performed using the hypergeometric distribution (one-tailed). Gene sets with fewer than five genes after filtering for expressed genes were removed before the gene set enrichment false discovery rates were calculated.

To calculate the statistics for overlap of *daf-2* upregulated genes, *spr-4;spr-3;daf-2* upregulated genes, *spr-4;spr-3;daf-2* downregulated genes, and *daf-16* class I genes for *C. elegans* day 10 we used the SuperExactTest version 1.0.0 R package[60].

For meta-analysis of GO terms and ENCODE ChIP–seq transcription factor target gene set enrichment from the ROSMAP, CMC, and Gibbs cohorts, we used Stouffer's method with weights[61] for combining $P$ values implemented in the sumz function in the metap R package[62]. $P$ values equal to 1 were replaced with 0.999999 to comply with the requirement that $0 < P < 1$. For study weights we used the square root of the total number of individuals in the combined youngold and oldestold group for each study (ROSMAP, sqrt(117); CMC, sqrt(155); Gibbs, sqrt(37))[61].

**Hierarchical clustering to select age group cutoffs.** Normalized and adjusted gene expression values were derived as described in the section Gene expression normalization and covariate adjustment, with the following changes. First, the group variable was a factor with levels cogdx1, cogdx2, cogdx3, cogdx4, cogdx5, and cogdx6 (ROSMAP); control, schizophrena, bipolar, affective (CMC); or not included (Gibbs). Next, genes were deemed expressed if ≥10 individuals in the cognitively normal group (ROSMAP, cogdx1; CMC, control with age ≥ 60) had more than 1 CPM. Genes not satisfying these criteria were removed, keeping the original library sizes. For the Gibbs cohort, probes were deemed expressed if ≥10 individuals with age ≥ 55 had detection $P < 0.01$ and no samples had NA values. Probes not satisfying these criteria were removed. These were expressed genes used for analysis. This filtering retained: ROSMAP, 18,734 expressed genes; CMC, 19,615 expressed genes; Gibbs, 13,411 expressed probes. For RNA-seq data, counts were then normalized using the TMM method in edgeR, and log(CPM) values were calculated for expression. Finally, age was included as a continuous covariate, and the effects of group and age were added back to the residuals. Then cognitively normal aged individuals were selected in each cohort (ROSMAP, cogdx1; CMC, control with age ≥ 60; Gibbs, age ≥ 55). To determine age-associated genes, the expression of each expressed gene was correlated with the age of death of the individual using Spearman's correlation. Genes with Spearman rank correlation FDR < 0.1 (ROSMAP, $n = 1{,}025$ genes) or FDR < 0.05 (CMC, $n = 6{,}828$ genes; Gibbs, $n = 203$ probes) were age-associated genes. Permuting sample ages before calculating the Spearman correlations gave no genes that passed the FDR threshold. Genes were then normalized to have mean 0 and s.d. 1 across individuals, and genes with normalized expression ≥3 or ≤−3 were set to 3 or −3, respectively. The Pearson correlation coefficient between individuals was then calculated using only age-associated genes.

Agglomerative hierarchical clustering using the Euclidean distance metric and average linkage criterion was performed on the matrix of pair-wise Pearson correlations to cluster individuals. The resulting tree was cut into three groups, and the distribution of the age of death of individuals of each group was analysed. Using the ROSMAP cohort, for the cluster with the smallest median age, with c1 individuals, we calculated the number of individuals, nyoc, with age of death in [70, youb], of nyo total, and for the cluster with the largest median age, of c2 individuals, we calculated the number of individuals, nooc, with age of death in [oolb, Inf), of noo total, as the age boundaries youb and oolb varied. For each set of age cutoffs we assigned a score as (nyoc/nyo × c1 + nooc/noo × c2)/(c1 + c2). To ensure adequate sample sizes for differential expression and because individuals with older ages appeared more heterogeneous in the cluster heat map, we considered only age cutoffs with at least 25 individuals in the youngold group and at least 75 individuals in oldestold group. On the basis of the top scoring age cutoffs in the ROSMAP cohort, we selected youb = 80 and oolb = 85. Thus, the age cutoffs for defining the youngold, middleold, and oldestold groups used for differential expression in the ROSMAP, CMC, and Gibbs cohorts were selected to be ≤80, >80 and <85, and ≥85, respectively.

**Statistical analysis and data representation.** Statistical analysis was performed using R. Statistical tests used are noted in the figure legends or in the relevant Methods section. Throughout the paper, all tests are two-sided and unpaired unless stated otherwise. A significance level of 0.05 was used to reject the null hypothesis unless stated otherwise.

Box plots throughout the paper show the median, lower and upper hinges (first and third quartiles), upper whisker (hinge to the largest value no further than 1.5 × IQR from the hinge), lower whisker (hinge to the smallest value at most 1.5 × IQR from the hinge), and outlying points beyond the whiskers. Additionally, all points are plotted on top of the box plot and randomly jittered horizontally

A $t$-test was used for parametric comparisons between two groups with normally distributed data, and used t.test(). For groups with equal variance, Student's $t$-test was used (argument var.equal = TRUE); otherwise Welch's $t$-test was used (argument var.equal = FALSE). ANOVA was used for parametric comparisons between more than two groups and used aov(). The Mann–Whitney $U$-test was used for nonparametric comparisons between two groups without knowledge of their distribution and used wilcox.test() with arguments exact = TRUE, correct = FALSE. Levene's test was used to investigate the homogeneity of variance across groups and used leveneTest() from the 'car' R package. $Q$–$Q$ plots and the Shapiro–Wilk test were used to assess normality and used shapiro.test(). The Bonferroni outlier test was used to assess outliers and used outlierTest() from the 'car' R package. The log-rank test was used to compare the survival distributions of two groups. Survival statistics were calculated using survfit() with argument type = 'kaplan-meier' and survdiff(). To determine whether there was a linear relationship between two variables we fit a linear regression model and tested the null hypothesis that the slope of the regression line is 0 using a two-sided $t$-test using lm(). $P$ values were corrected for multiple comparisons where noted using p.adjust() with argument method = 'fdr', 'holm', or 'bonferroni' for false discovery rate, Holm's method, or Bonferroni correction, respectively. The GCaMP imaging data indicated that nonparametric analyses were most appropriate and thus were used for all GCaMP analyses. In Fig. 2e, the presence of candidate outliers and/or potential non-normality for time points 1–12.5 min suggested that a Mann–Whitney $U$-test was more appropriate although a $t$-test produced similar conclusions. Meta-analysis of lifespan experiments that performed pairwise comparisons among more than two groups (Extended Data Fig. 8a right, Extended Data Fig. 8c) used pairwise Student's $t$-tests implemented in pairwise.t.test() with argument pool.sd = TRUE to calculate a common s.d. used for all groups and comparisons.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Data from the ROSMAP cohort are available under controlled use conditions set by human privacy regulations. To access the data, a data use agreement is needed. This registration is in place solely to ensure the anonymity of the ROSMAP study participants. Data can be requested on the Rush Alzheimer's Disease Center Resource Sharing Hub at http://www.radc.rush.edu/. *C. elegans* RNA-seq data are available in the Gene Expression Omnibus (GEO) under accession number GSE123146.

## Code availability

All code used in the analysis of data presented in this manuscript is available upon request.

29. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9**, 628–645 (2012).
30. Bennett, D. A. et al. Overview and findings from the Rush Memory and Aging Project. *Curr. Alzheimer Res.* **9**, 646–663 (2012).
31. Mu, X. et al. A gene network downstream of transcription factor Math5 regulates retinal progenitor cell competence and ganglion cell fate. *Dev. Biol.* **280**, 467–481 (2005).
32. Mao, C. A. et al. Neuronal transcriptional repressor REST suppresses an Atoh7-independent program for initiating retinal ganglion cell development. *Dev. Biol.* **349**, 90–99 (2011).
33. Dhamne, S. C. et al. Replicable in vivo physiological and behavioral phenotypes of the *Shank3B* null mutant mouse model of autism. *Mol. Autism* **8**, 26 (2017).
34. Kelly, E. et al. mGluR5 modulation of behavioral and epileptic phenotypes in a mouse model of tuberous sclerosis complex. *Neuropsychopharmacology* **43**, 1457–1465 (2018).
35. Purtell, H. et al. Electrographic spikes are common in wildtype mice. *Epilepsy Behav.* **89**, 94–98 (2018).
36. Yuskaitis, C. J. et al. A mouse model of DEPDC5-related epilepsy: neuronal loss of Depdc5 causes dysplastic and ectopic neurons, increased mTOR signaling, and seizure susceptibility. *Neurobiol. Dis.* **111**, 91–101 (2018).
37. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
38. Lakowski, B. et al. Two suppressors of sel-12 encode C2H2 zinc-finger proteins that regulate presenilin transcription in *Caenorhabditis elegans*. *Development* **130**, 2117–2128 (2003).
39. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
40. Ohnishi, N., Kuhara, A., Nakamura, F., Okochi, Y. & Mori, I. Bidirectional regulation of thermotaxis by glutamate transmissions in *Caenorhabditis elegans*. *EMBO J.* **30**, 1376–1388 (2011).
41. Jarriault, S. & Greenwald, I. Suppressors of the egg-laying defective phenotype of *sel-12* presenilin mutants implicate the CoREST corepressor complex in LIN-12/Notch signaling in *C. elegans*. *Genes Dev.* **16**, 2713–2728 (2002).
42. Winston, W. M., Molodowitch, C. & Hunter, C. P. Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. *Science* **295**, 2456–2459 (2002).
43. *C. elegans* Deletion Mutant Consortium. Large-scale screening for targeted knockouts in the *Caenorhabditis elegans* genome. *G3 (Bethesda)* **2**, 1415–1425 (2012).
44. Nottke, A. C. et al. SPR-5 is a histone H3K4 demethylase with a role in meiotic double-strand break repair. *Proc. Natl Acad. Sci. USA* **108**, 12805–12810 (2011).
45. Greer, E. L., Becker, B., Latza, C., Antebi, A. & Shi, Y. Mutation of *C. elegans* demethylase spr-5 extends transgenerational longevity. *Cell Res.* **26**, 229–238 (2016).

46. Robert, V. J., Katic, I. & Bessereau, J. L. Mos1 transposition as a tool to engineer the *Caenorhabditis elegans* genome by homologous recombination. *Methods* **49**, 263–269 (2009).
47. Kamath, R. S. & Ahringer, J. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* **30**, 313–321 (2003).
48. Frøkjær-Jensen, C., Davis, M. W., Ailion, M. & Jorgensen, E. M. Improved Mos1-mediated transgenesis in *C. elegans*. *Nat. Methods* **9**, 117–118 (2012).
49. Sarov, M. et al. A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. *Cell* **150**, 855–866 (2012).
50. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
52. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
54. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
55. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
56. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
57. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
58. Tepper, R. G. et al. PQM-1 complements DAF-16 as a key transcriptional regulator of DAF-2-mediated development and longevity. *Cell* **154**, 676–690 (2013).
59. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology. *R package version* **2340** (2018).
60. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 16923 (2015).
61. Zaykin, D. V. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* **24**, 1836–1841 (2011).
62. Dewey, M. Metap: meta analysis of significance values. R package Version 10 (2018).

**Acknowledgements** We thank members of the Yankner laboratory for suggestions and discussion, and the Hsieh laboratory for providing the anti-mouse REST antibody. This work was supported by an NIH Director's Pioneer Award (DP1OD006849) and NIH grants RO1AG046174 and RO1AG26651 to B.A.Y., RO1GM072551 to M.P.C., P30AG10161, R01AG15819, R01AG17917, R01AG36836 and U01AG46152 to D.A.B., EY024376 to C.-A.M., EY011930 to W.H.K. and K99AG050830 to J.M.Z., and grants from the Glenn Foundation for Medical Research and The Ludwig Family Foundation to B.A.Y. Data were generated as part of the CMC, supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881, R37MH057881S1, HHSN271201300031C, AG02219, AG05138 and MH06692.

**Author contributions** J.M.Z., P.O. and N.D. performed experiments in *C. elegans*; D.D. performed statistical analysis and informatics on human brain and *C. elegans* RNA-seq; L.A. and D.D. analysed *Rest* cKO mice and human brain sections; L.A., S.C.D. and A.R. performed PET–CT and electrophysiological analysis of mice; J.M.Z. performed cell culture; C.-A.M., W.H.K. and G.M.C. contributed reagents; D.A.B. contributed brain samples and data from the ROSMAP cohort; J.M.Z., D.D., M.P.C. and B.A.Y. contributed to the overall study design; B.A.Y. directed the study; and B.A.Y., J.M.Z. and D.D. wrote the manuscript, which was examined by all authors.

**Competing interests** G.M.C. is a cofounder and senior advisor for GC Therapeutics, Inc, which uses transcription factors for therapeutics. The other authors declare no competing interests.

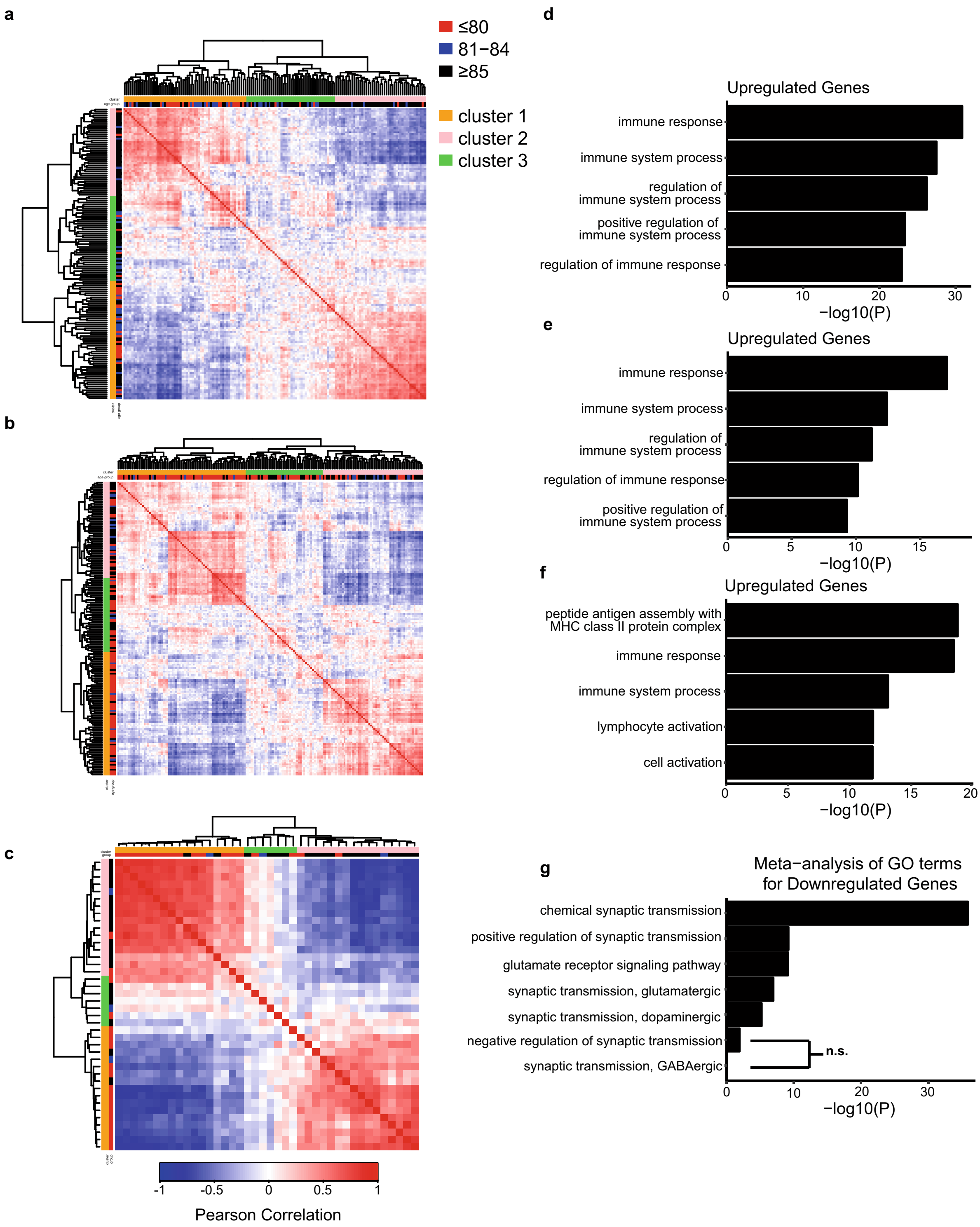
**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1647-8.

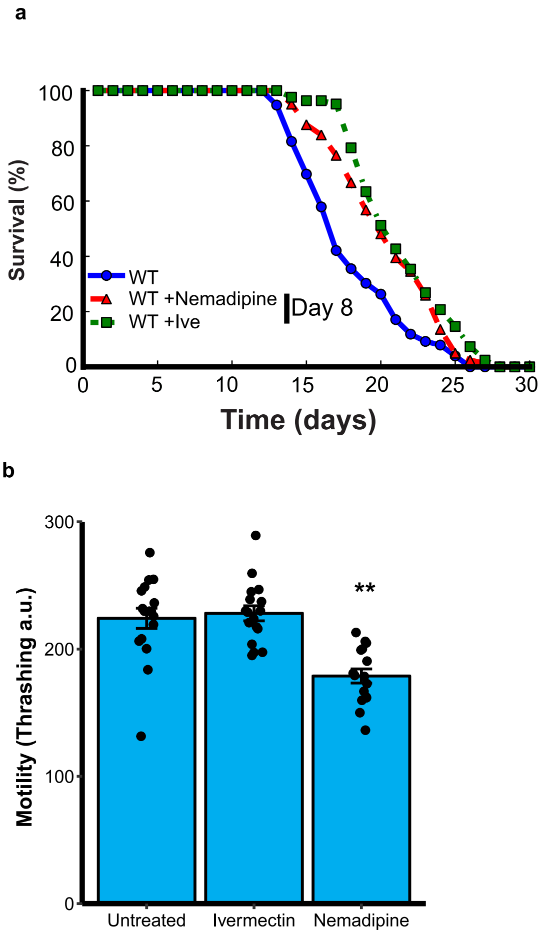
**Correspondence and requests for materials** should be addressed to B.A.Y.

**Peer review information** *Nature* thanks Nektarios Tavernarakis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

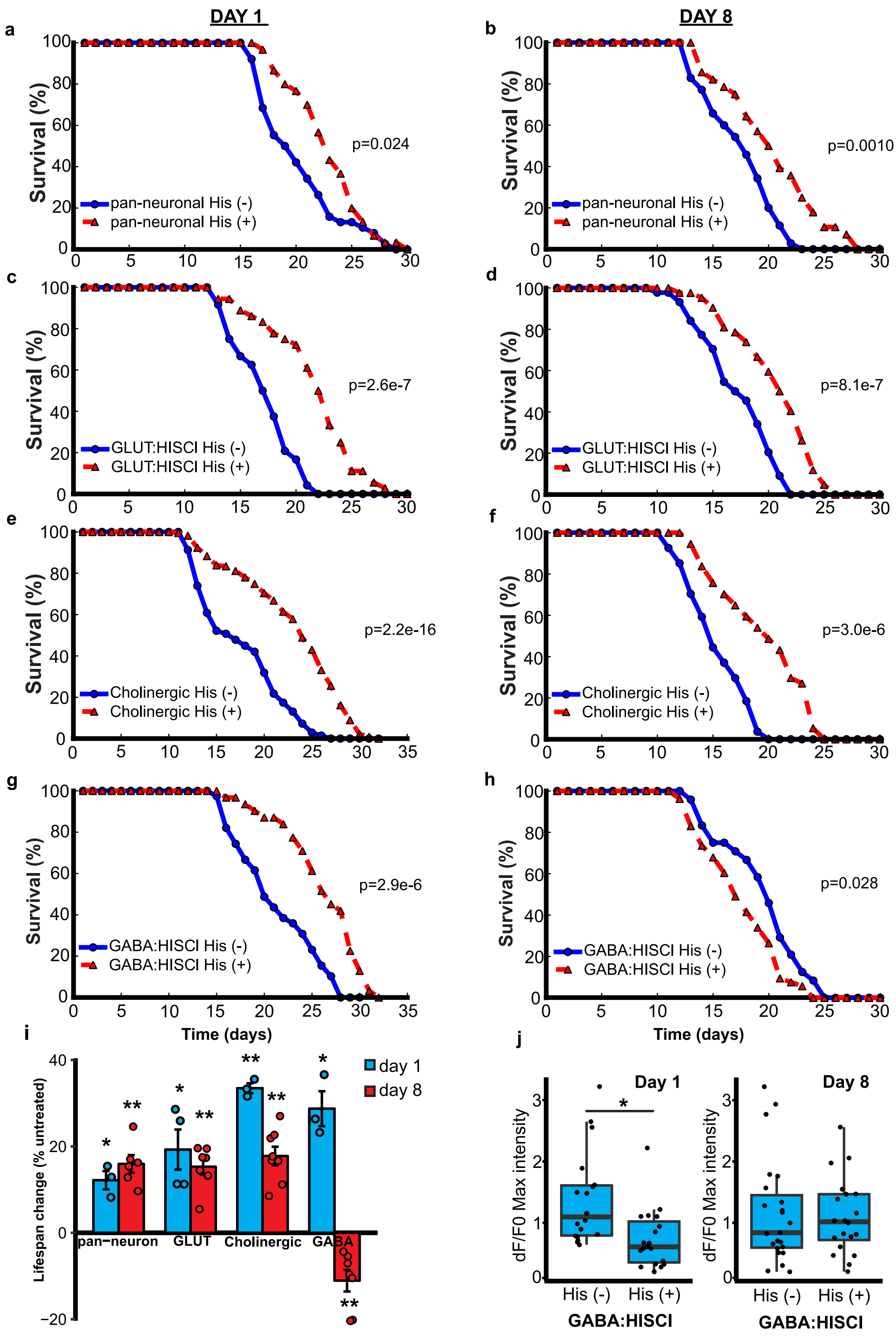
**Reprints and permissions information** is available at http://www.nature.com/reprints.



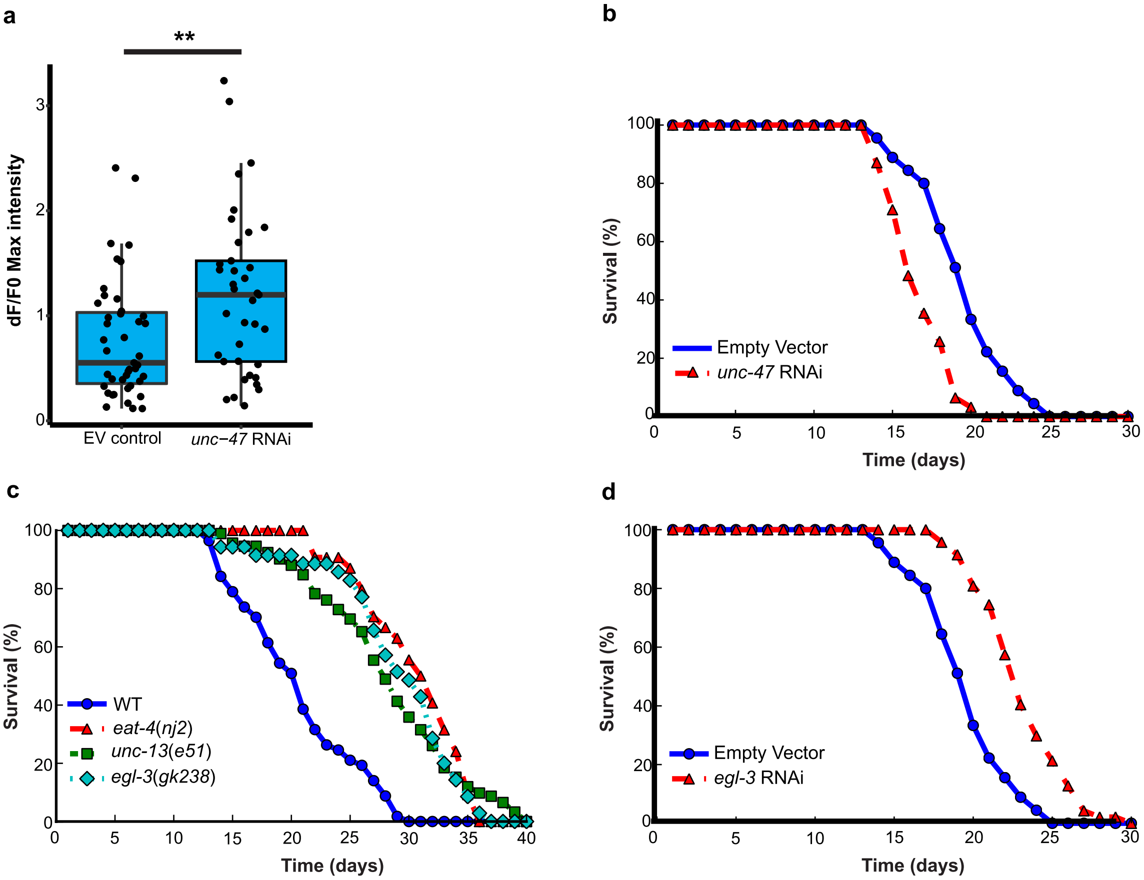
**Extended Data Fig. 1 | Partitioning of the ageing human population for analysis of gene expression in the brain. a–c**, Adjusted gene expression profiles of age-associated genes were compared between cognitively normal aged individuals to derive a matrix of Pearson correlation coefficients that indicate the degree of similarity between any two cases in the ROSMAP (**a**, dorsolateral prefrontal cortex, $n = 150$ individuals), CMC (**b**, dorsolateral prefrontal cortex, $n = 174$ individuals) and Gibbs (**c**, frontal cortex, $n = 40$ individuals) cohorts. **d–f**, Most significantly enriched GO terms for upregulated genes in the cortex of cognitively normal individuals who lived to be ≥85 years old relative to individuals who lived to be ≤80 years old in the ROSMAP (**d**, $n = 117$ individuals), CMC (**e**, $n = 155$ individuals), and Gibbs (**f**, $n = 37$ individuals) cohorts. $P$ values were calculated using Fisher's exact test (see Methods). **g**, Meta-analysis of GO term enrichment for downregulated genes. Shown are selected GO terms related to excitatory and inhibitory synaptic transmission. The individual cohort enrichment $P$ values were combined using Stouffer's method (see Methods). NS, not significant (FDR > 0.1).



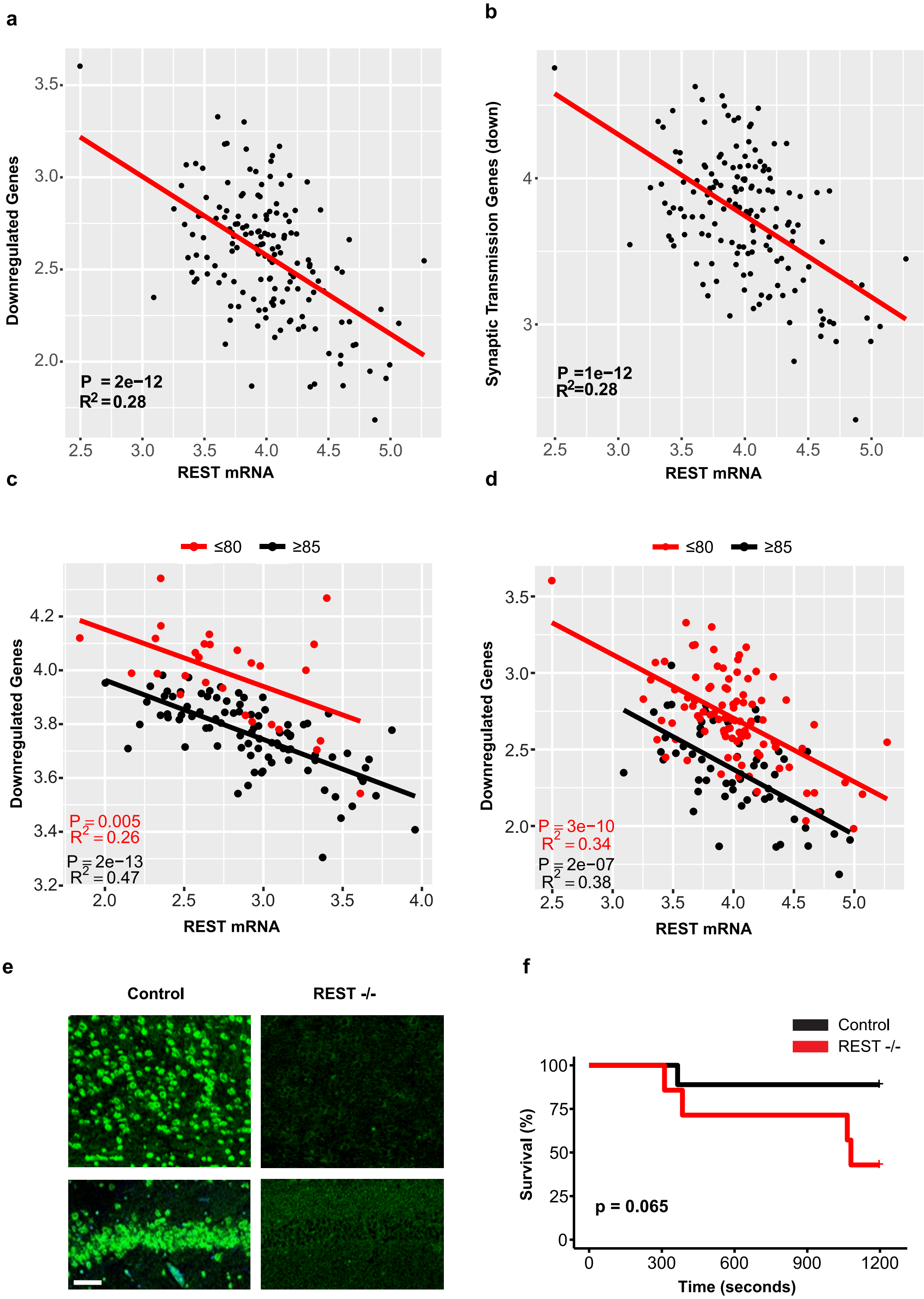
**Extended Data Fig. 2 | Ivermectin and nemadipine extend lifespan without interfering with worm motility. a**, Worms were transferred at day 8 to either standard NGM plates or plates containing ivermectin (Ive, 1 pg ml$^{-1}$) or nemadipine (2 μM). Shown is a representative curve of an experiment repeated twice. Nemadipine versus wild-type, $P = 3.2 \times 10^{-4}$; ivermectin versus wild-type, $P = 2.2 \times 10^{-7}$ by log-rank test. Nemadipine, $n = 81$; ivermectin, $n = 82$; wild-type, $n = 76$. **b**, Day 2 worms treated with nemadipine or ivermectin for 24 h were transferred to liquid culture and thrashing rate was assessed using the Nemametrix wMicrotracker (see Methods). Shown are mean motility scores for the first 60 min ± s.e.m. Untreated, $n = 17$ wells; ivermectin, $n = 17$ wells; nemadipine, $n = 16$ wells. Each well contained about 10 worms. $^{**}P = 1.7 \times 10^{-4}$ versus untreated, Mann–Whitney *U*-test with multiple testing correction by Holm's method. Results are representative of an experiment replicated twice.



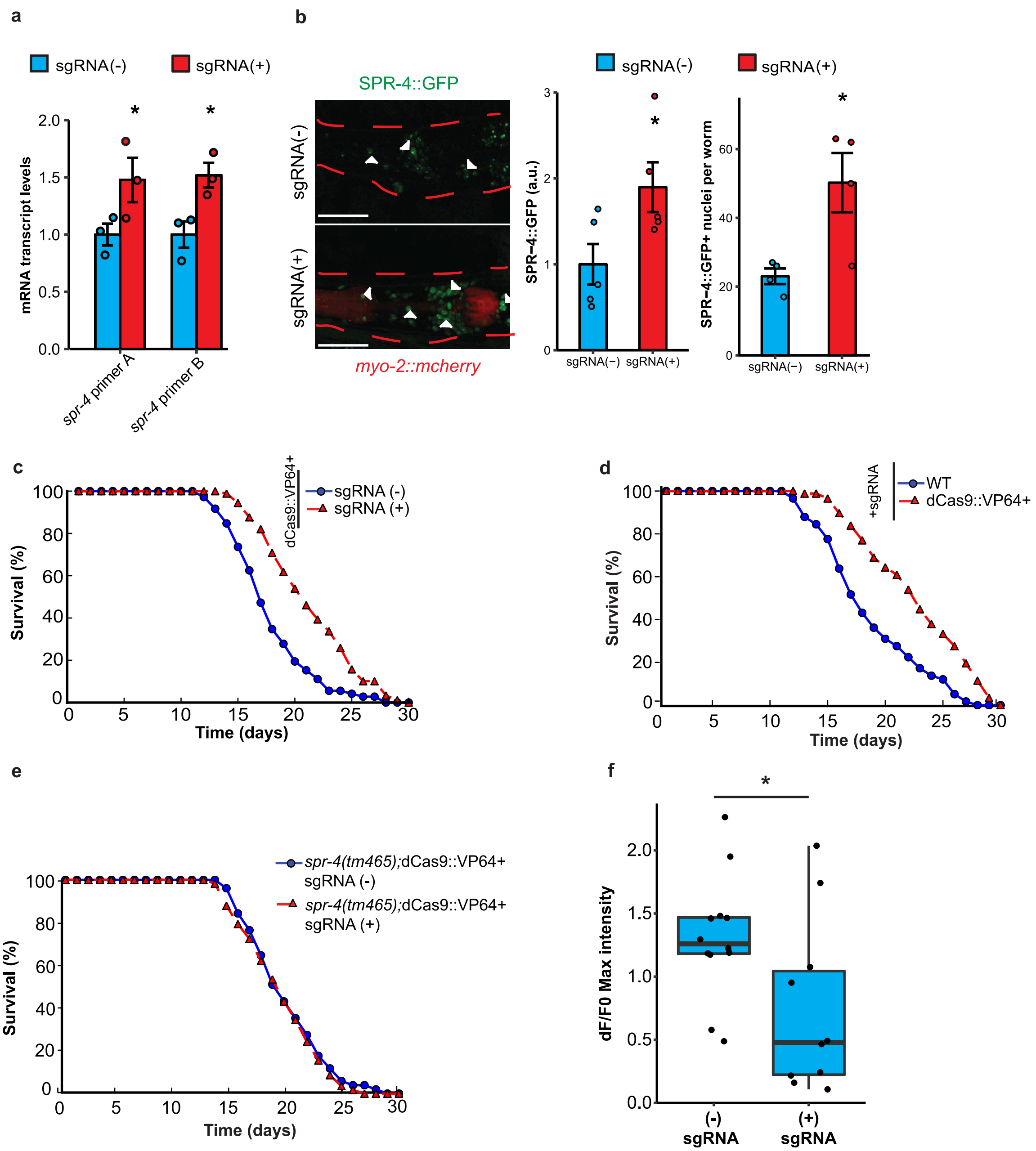
**Extended Data Fig. 3 | Repression of multiple neurotransmitter systems extends lifespan in *C. elegans*.** **a**–**h**, *C. elegans* lines expressing the transgenic HisCl1 channel in the indicated neuronal populations were treated with 10 mM histamine (His+) starting at adult day 1 (**a**, **c**, **e**, **g**) or day 8 (**b**, **d**, **f**, **h**) and compared to untreated controls (His−). *P* values calculated by log-rank test. See Supplementary Table 22 for individual *n* values and statistics. **i**, Mean lifespan extension ± s.e.m. for worms treated with histamine at days 1 or 8 relative to untreated controls for at least three independent replicates. $*P < 0.05$, $**P < 0.01$ by Student's *t*-test. HisCl1was driven using the GAL4SK:VP64 system for the GABAergic (GABA), glutamatergic (GLUT) and cholinergic systems, using *unc-47*, *eat-4*, and *unc-17* drivers, respectively (see Supplementary Table 19). **j**, Reduced ASH neuron excitation following inhibition of GABA activity at day 1 but not day 8. Shown is normalized maximum GCaMP fluorescence in day 1 and 8 *unc-47*:HisCl1 worms that were treated with 10 mM histamine (His (+)) on the indicated day, or untreated controls (His (−)). Day 1 His(−), $n = 18$ worms; day 1 His(+), $n = 19$ worms; day 8 His(−), $n = 23$ worms; day 8 His(+), $n = 20$ worms. $*P = 1.1 \times 10^{-3}$ by Mann–Whitney *U*-test.



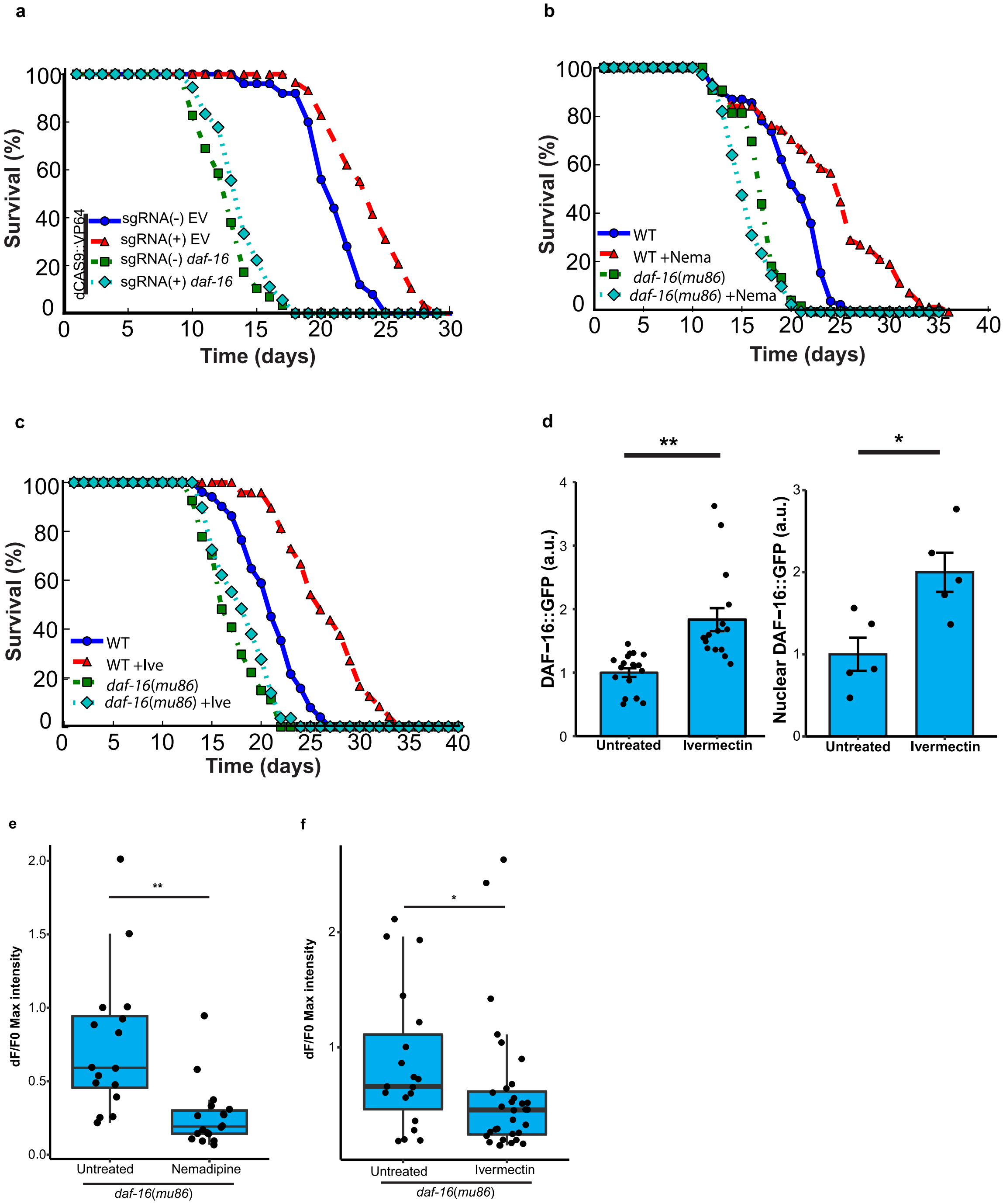
**Extended Data Fig. 4 | Neural excitation, neuropeptide signalling and lifespan in *C. elegans*.** **a**, Increased excitation of ASH neurons following RNAi against the GABA vesicular transporter *unc-47*. GCaMP imaging was performed on worms with enhanced neuronal RNAi (see Fig. 3 legend and Methods) for *unc-47* ($n = 37$) or controls ($n = 43$) at day 2. $^{**}P = 6.8 \times 10^{-3}$ by Mann–Whitney *U*-test. **b**, RNAi of *unc-47* reduces lifespan. Worms with enhanced neuronal RNAi were treated with *unc-47* ($n = 31$) or control RNAi ($n = 84$). Shown is a representative lifespan analysis replicated three times. $P = 1.3 \times 10^{-6}$ by log-rank test. **c**, Reduction in synaptic neurotransmission or neuropeptide signalling extends lifespan in *C. elegans*. Mutations in genes affecting glutamatergic neurotransmission (*eat-4*), presynaptic function (*unc-13*) and neuropeptide signalling (*egl-3)* produce comparable lifespan extensions. WT, $n = 57$; *eat-4*(*nj2*), $n = 54$, $P \leq 2.2 \times 10^{-16}$; *unc-13*(*e51*), $n = 92$, $P = 3.6 \times 10^{-14}$; *egl-3*(*gk238*), $n = 35$, $P = 8.3 \times 10^{-11}$ by log-rank test. Curves are representative of two independent replicates. **d**, Extension of lifespan by *egl-3* RNAi in worms with enhanced neuronal RNAi. Shown are lifespan curves representative of two independent replicates. *egl-3* RNAi ($n = 47$ worms); empty vector ($n = 84$ worms). $P = 3.5 \times 10^{-11}$ by log-rank test.



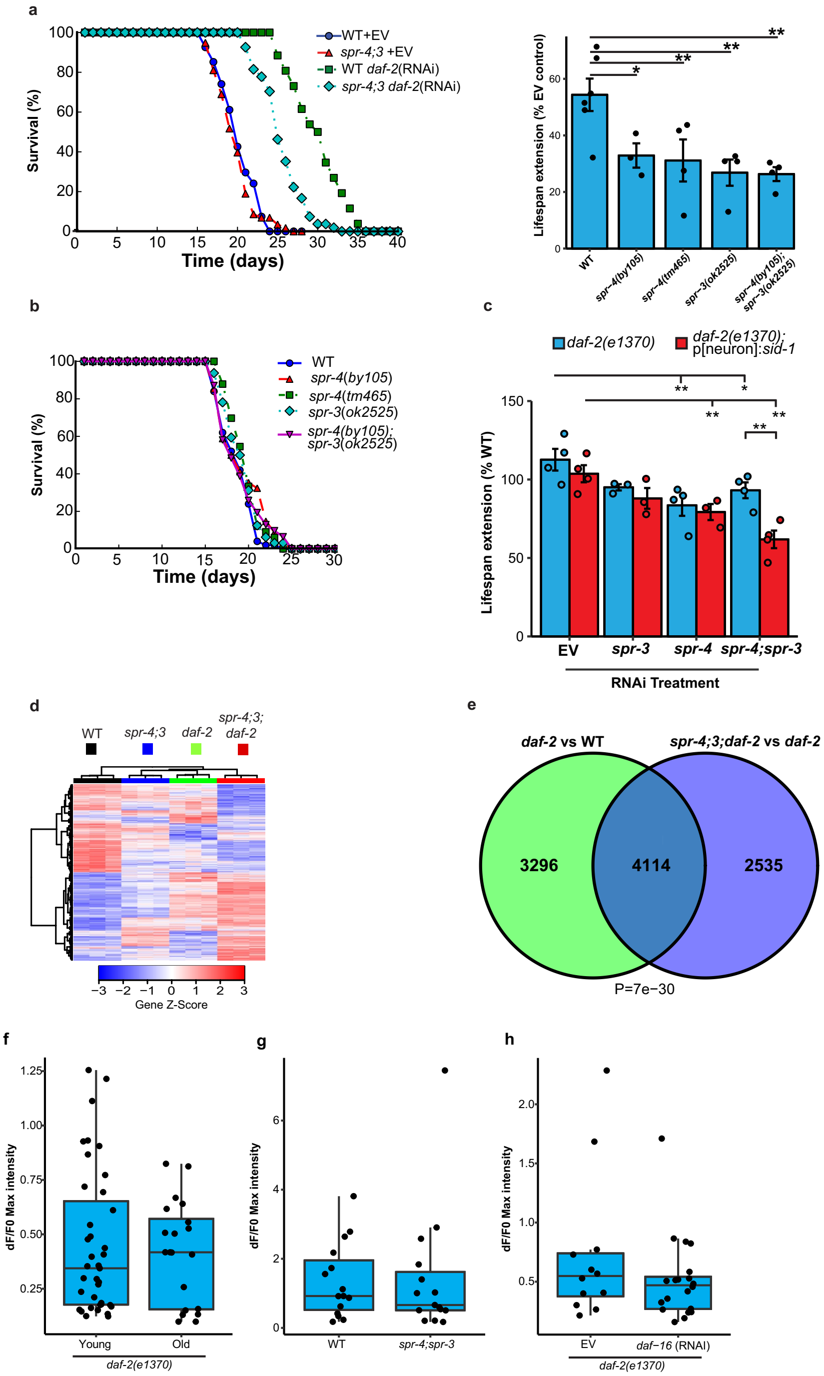
**Extended Data Fig. 5 | Gene regulation and neural activity associated with REST and extended longevity.** **a**, **b**, Expression of genes that are downregulated in individuals ≥85 years old versus ≤80 years old is inversely related to *REST* mRNA levels. Shown is linear regression analysis of normalized and adjusted *REST* mRNA levels and mean expression of all downregulated genes (**a**) and downregulated genes associated with the synaptic transmission GO term (**b**). Data are from the CMC cohort. Each point represents an individual case, $n = 155$ individuals. *P* values derived by a *t*-test for the slope of the regression line. Note similarity to the data for the ROSMAP cohort in Fig. 2a, b. **c**, **d**, Stratification by age group. Analysis of the ROSMAP cohort (**c**, $n = 117$ individuals) and the CMC cohort (**d**, $n = 155$ individuals) as in Fig. 2a, but stratified by age group. *P* values derived by *t*-test for the slope of the regression line. **e**, Loss of REST expression in conditional *Rest* knockout mice. Representative images of the cortex (top) and hippocampus (bottom) from $Rest^{lx/lx}$ (control) and *Nestin-Cre;*$Rest^{lx/lx}$ ($Rest^{-/-}$) mice. Immunolabelling was performed with the anti-mouse REST-14 antibody directed against the REST C-terminal domain (Supplementary Table 20). Scale bar, 40 μm. Image is representative of an experiment replicated four times. **f**, Survival of $Rest^{-/-}$ and control mice following administration of the seizure-inducing agent PTZ (40 mg $kg^{-1x}$r $Rest^{-/-}$ versus control by log-rank test. Control, $n = 9$; $Rest^{-/-}$, $n = 7$.



**Extended Data Fig. 6 | Induction of *spr-4* extends lifespan and suppresses neural excitation in *C. elegans*.** **a**, *spr-4* mRNA levels in worms expressing a stably integrated dCas9::VP64 transgene in the presence (sgRNA(+)) or absence (sgRNA(−)) of four different sgRNAs targeting the *spr-4* promoter. Transcript levels were determined by qRT–PCR and normalized to sgRNA(−) controls. Mean ± s.e.m., $n = 3$. Primer A, $*P = 0.041$; primer B, $*P = 0.020$ by one-sided Student's *t*-test. **b**, dCas9::VP64-mediated elevation of SPR-4 protein levels. Left, representative images of the head region of heterozygous $F_1$ progeny of the strains bearing a *pspr-4::spr-4::gfp::spr-4utr* transgene. Arrowheads indicate SPR-4::GFP-positive nuclei. Dashed red lines indicate the outline of the worm body. Scale bar, 40 μm. Middle, SPR-4::GFP protein levels are increased by dCAS9::VP64 mediated activation. Values represent mean ± s.e.m. sgRNA(−), $n = 5$ worms; sgRNA(+), $n = 5$ worms with 7–38 measurements per worm; $*P = 0.022$, one-sided Student's *t*-test. Right, SPR-4::GFP expression appears in more cells following dCAS9::VP64-mediated activation. Values represent mean ± s.e.m. sgRNA(−), $n = 4$ worms; sgRNA(+), $n = 4$ worms. $*P = 0.011$, one-sided Student's *t*-test. Shown is a representative experiment replicated three times. **c**, Extended lifespan in worms expressing an integrated *dCas9::VP64* transgene and sgRNAs targeting the *spr-4* promoter (sgRNA(+)) ($n = 79$ worms) relative to dCas9::VP64-expressing worms in the absence of sgRNAs (sgRNA(−)) ($n = 57$ worms). $P = 5.5 \times 10^{-9}$, log-rank test. Representative of an experiment replicated six times. **d**, Lifespans of worms expressing sgRNA targeting the *spr-4* promoter in the presence ($n = 87$ worms) or absence ($n = 58$ worms) of dCas9::VP64. $P = 3.7 \times 10^{-7}$, log-rank test. Representative of an experiment replicated twice. **e**, Lifespans of dCas9::VP64-expressing worms in the presence ($n = 51$ worms) or absence ($n = 58$ worms) of sgRNAs on the *spr-4(tm465)* loss-of-function mutant background. $P = 0.49$, log-rank test. Representative of three independent replicates. **f**, Overexpression of *spr-4* reduces neural excitation. GCaMP imaging was performed in ASH neurons in SPR-4-overexpressing (sgRNA(+)) and control (sgRNA(−)) worms in the lines described in **c**. Shown are maximum GCaMP fluorescence changes. sgRNA(−), $n = 12$ worms; sgRNA(+), $n = 10$ worms. $*P = 0.025$, Mann–Whitney *U*-test.



**Extended Data Fig. 7 | Lifespan extension by overexpression of *spr-4* and inhibition of neural excitation depends on DAF-16. a**, Extension of lifespan by overexpression of *spr-4* is dependent on DAF-16. Lifespans of worms overexpressing *spr-4* (sgRNA(+); dCAS9::VP64) or not overexpressing *spr-4* (sgRNA(−); dCAS9::VP64) following treatment with *daf-16* RNAi or an empty vector control. sgRNA(+) EV ($n = 29$ worms) versus sgRNA(−) EV ($n = 25$ worms): $P = 2.7 \times 10^{-4}$; sgRNA(+) *daf-16* ($n = 18$ worms) versus sgRNA(−) *daf-16* ($n = 29$ worms): $P = 0.20$ by log-rank test. Representative of four independent replicates. **b**, **c**, Extension of lifespan by the inhibitors of neural excitation ivermectin and nemadipine is DAF-16-dependent. Shown are lifespan determinations for wild-type control and *daf-16*(*mu86*) mutant worms in the presence or absence of nemadipine (2 μM; **b**) or ivermectin (1 pg ml$^{-1}$; **c**). **b**, WT, $n = 69$ worms; WT + Nema, $n = 51$; *daf-16*, $n = 43$; *daf-16* + Nema, $n = 67$. WT + Nema versus WT, $P = 9.9 \times 10^{-8}$; *daf-16* + Nema versus *daf-16*, $P = 0.014$; log-rank test. **c**, WT, $n = 78$ worms; WT + Ive, $n = 77$; *daf-16*, $n = 27$; *daf-16* + Ive, $n = 29$. WT + Ive versus WT, $P = 7.3 \times 10^{-8}$; *daf-16* + Ive versus *daf-16*, $P = 0.22$; log-rank test. Curves are representative of an experiment replicated two (nemadipine) or three (ivermectin) times. **d**, Inhibition of neural excitation with ivermectin elevates DAF-16 levels. Worms expressing a *Daf-16::GFP* transgene were treated for 10 days with 1 pg ml$^{-1}$ ivermectin and assessed by confocal microscopy. Left, total DAF-16::GFP (mean ± s.e.m.). Untreated, $n = 19$ worms; ivermectin, $n = 16$ worms. $^{**}P = 2.5 \times 10^{-7}$, Mann–Whitney *U*-test. Right, nuclear DAF-16::GFP. $n = 5$ worms per group, 50–61 nuclei per worm. $^{*}P = 0.013$ by Student’s *t*-test. Results are representative of an experiment replicated twice. **e**, DAF-16 is not required for inhibition of neural excitation by nemadipine. Shown are maximum ASH GCaMP intensity changes for day 2 *daf-16*(*mu86*) mutant worms treated for 24 h with 2 μM nemadipine (untreated, $n = 16$ worms; nemadipine, $n = 18$ worms). $P = 9.4 \times 10^{-5}$, Mann–Whitney *U*-test. **f**, DAF-16 is not required for inhibition of neural excitation by ivermectin. Shown are data from day 2 worms treated for 24 h with 1 pg ml$^{-1}$ ivermectin (control, $n = 19$ worms; ivermectin, $n = 32$ worms). $P = 0.030$, Mann–Whitney *U*-test.

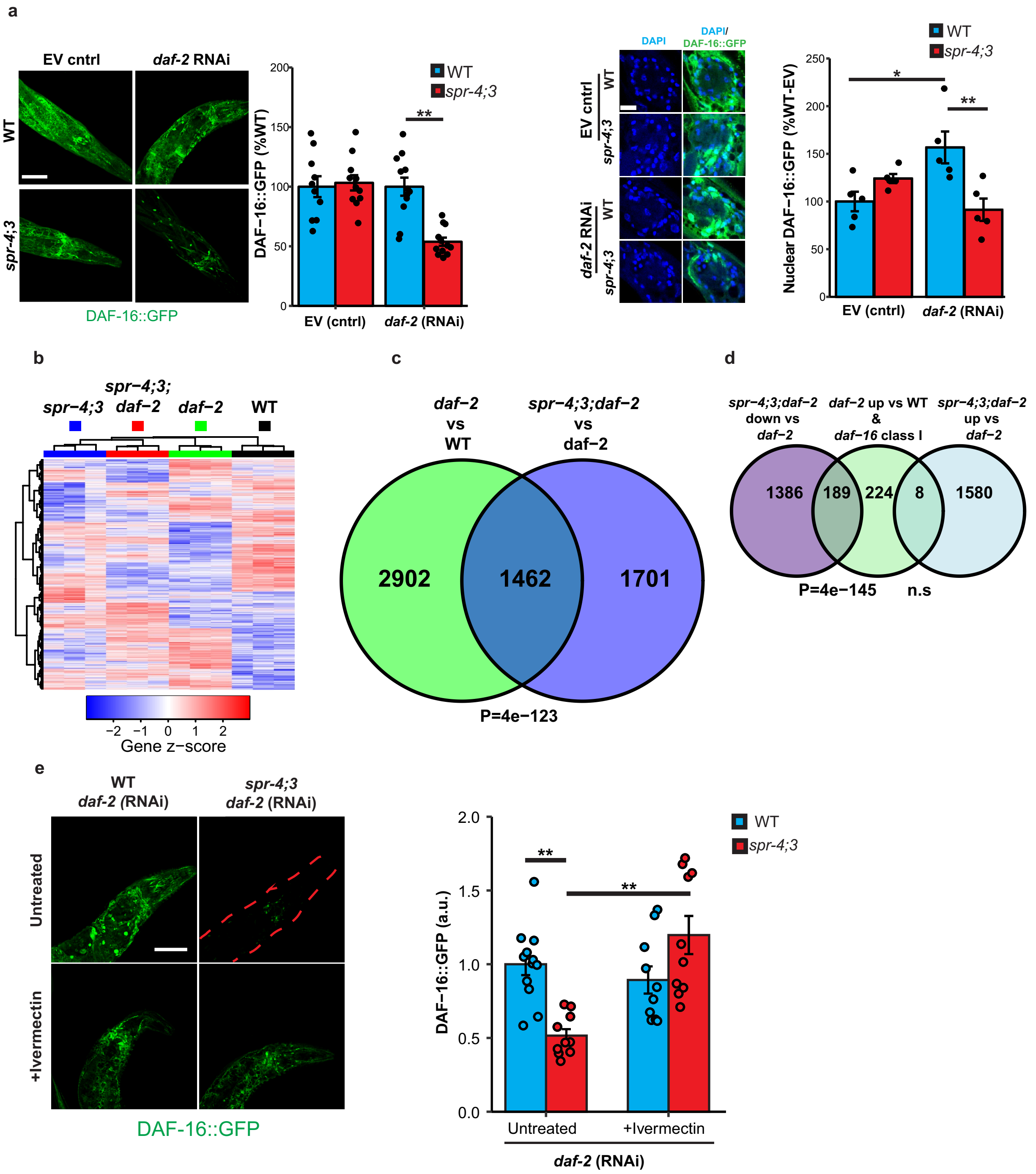


i

| Gene name | Function |
|---|---|
| *eat-4* | Glutamate vesicular transporter |
| *unc-47* | GABA vesicular transporter |
| *cat-2* | Tyrosine hydroxylase (catacholoamine synthesis) |
| *cha-1/unc-17* | Acetylcholine synthase/Acetylcholine transporter |
| *tph-1* | Tryptophan hydroxylase |
| *egl-3* | Proprotein convertase (neuropeptide secretion) |
| *unc-104* | Axonal kinesin |
| *snb-1* | Synaptobrevin |
| *unc-13* | SNARE regulator |

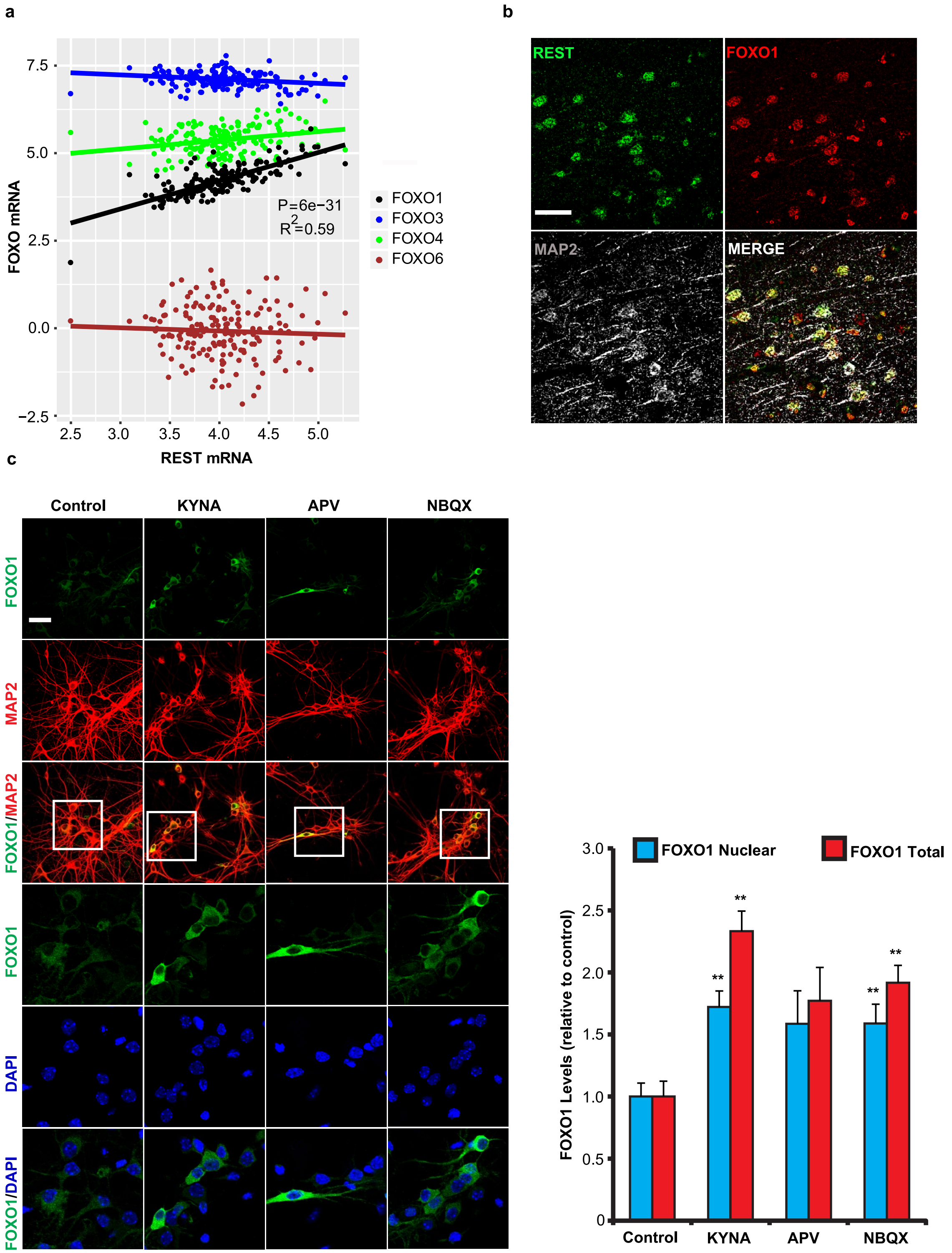
**Extended Data Fig. 8 |** See next page for caption.

**Extended Data Fig. 8 | SPR-3 and SPR-4 contribute to lifespan extension and gene regulation associated with reduced DAF-2 and insulin–IGF-like signalling. a**, Loss of function of SPR-3 and SPR-4 reduces the lifespan extension induced by *daf-2* RNAi. Left, representative lifespan analysis of *spr-4(by105);spr-3(ok2525)* double mutant and wild-type worms following *daf-2* or empty vector control RNAi. WT + EV, $n = 54$ worms; *spr-4;3* + EV, $n = 58$ worms; WT + *daf-2*, $n = 26$ worms; *spr-4;3* + *daf-2*, $n = 54$ worms. Right, values represent mean ± s.e.m. per cent lifespan extension (*daf-2* RNAi versus EV control) in the indicated genotypes. WT, $n = 6$ independent experiments; *spr-4*(*by105*), $n = 3$, $*P = 0.017$ versus WT; *spr-4*(*tm465*), $n = 4$, $**P = 0.0062$ versus WT; *spr-3*(*ok2525*), $n = 4$, $**P = 0.0018$ versus WT; *spr-4*(*by105*);*spr-3*(*ok2525*), $n = 4$, $**P = 0.0016$ versus WT; Students *t*-test. See Supplementary Table 22 for individual lifespan data and statistics. **b**, Lifespan is unaffected by *spr-4* and *spr-3* mutations in a wild-type background. WT, $n = 50$ worms; *spr-3*(*ok2525*), $n = 31$; *spr-4*(*by105*);*spr-3*(*ok2525*), $n = 32$; *spr-4*(*by105*), $n = 34$; *spr-4*(*tm465*), $n = 33$.There were no reproducibly significant changes by the log-rank test in 3–6 independent experiments per genotype (see Supplementary Table 22). **c**, Quantification of lifespan extension in *daf-2* mutant worms shown in Fig. 3b attributable to neuronal expression of *spr-3* and *spr*-4. RNAi was targeted to neurons by neuronal expression of a *sid-1* transgene in otherwise *sid-1*-null *daf-2(1370)* mutants (*daf-2*;p[neuron]:*sid-1*), and compared with untargeted RNAi in *sid-1* wild-type *daf-2(1370)* mutants (*daf-2*). Values represent mean ± s.e.m. lifespan extension relative to the control *sid-1*(*pk3321*);p[neuron]:*sid-1* worms treated with empty vector ($n = 3$ independent experiments). Significant lifespan effects were not observed for RNAi in the absence of the *daf-2* mutation. $*P < 0.05$; $**P < 0.01$ by Student's *t*-test. **d**, Gene expression determined by RNA-seq in day 2 adult worms. Differentially expressed genes (rows) and the indicated worm genotypes (columns) were clustered, and gene expression, transformed to a *z*-score per gene, is represented in a heat map. $n = 3$ independent replicates per genotype. **e**, Venn diagram illustrating the overlap in differentially expressed genes in *daf-2* single mutant versus WT and *spr-4;spr-3;daf-2* triple mutant versus *daf-2* single mutant worms. $P = 7 \times 10^{-30}$, Fisher's exact test with a one-sided alternative hypothesis. **f**, Long-lived *daf-2* mutants do not show an age-related increase in neural excitation. Shown are maximum ASH GCaMP intensity changes in day 1–2 ($n = 39$) and day 14–16 ($n = 20$) *daf-2(e1370)* mutant worms. Note the absence of the age-related increase in excitation observed in wild-type ageing worms (Fig. 1e). $P = 0.93$, Mann–Whitney *U*-test. **g**, The *spr-4;spr-3* double mutation in a wild-type background does not significantly affect neural excitation in ASH neurons. WT, $n = 15$ worms; *spr-4;spr-3*, $n = 15$ worms. $P = 0.62$, Mann–Whitney *U*-test. **h**, DAF-16 does not mediate suppression of neural excitation in the *daf-2* mutant. RNAi against *daf-16* was performed in *daf-2(e1370)* mutant worms on a *sid-1*(*pk3321*);p[neuron]:*sid-1* background to augment RNAi in neurons (*daf-16* RNAi, $n = 20$ worms, EV control, $n = 12$ worms). $P = 0.33$, Mann–Whitney *U*-test. **i**, Descriptions of the genes targeted by RNAi in Fig. 4d.



**Extended Data Fig. 9 |** See next page for caption.

**Extended Data Fig. 9 | Regulation of DAF-16 by SPR-3 and SPR-4.** **a**, Reduced DAF-16 activation in *spr-4;spr-3* mutants following *daf-2* R NAi. Left confocal image, day 10 worms of the indicated genotypes expressing an integrated *Daf-16::GFP* transgene and treated with *daf-2* RNAi or empty vector control since day 1 of adulthood. Images are maximum intensity *z*-projections. Scale bar, 40 µm. Left bar graph, mean ± s.e.m. GFP intensity in the peri-pharyngeal regions of *spr-4;spr-3* double mutants relative to wild-type controls for a representative experiment replicated four times (see Methods for details of analysis). $n = 8$–12 worms per replicate. $^{**}P = 5.2 \times 10^{-5}$, Welch's *t*-test. Right confocal image, higher-magnification views of DAF-16::GFP and DAPI-labelled nuclei. Images are magnified confocal *z*-planes. Scale bar, 10 µm. Right bar graph, mean ± s.e.m. nuclear GFP intensity relative to the WT-EV control, $n = 5$ worms per genotype and 51–89 nuclei per worm. $^{*}P = 0.016$, $^{**}P = 5.5 \times 10^{-3}$ by ANOVA with post hoc Tukey test. Values and images are representative of an experiment replicated three times. **b**, Gene expression determined by RNA-seq in adult day 10 worms. Differentially expressed genes (rows) and replicates of the indicated worm genotypes (columns) were clustered, and gene expression, transformed into a *z*-score per gene, is represented in a heat map. $n = 3$ independent replicates per genotype. **c**, Venn diagram illustrating the overlap of differentially expressed genes in day 10 *daf-2* versus wild-type and *spr-4;spr-3;daf-2* versus *daf-2* worms. $P = 4 \times 10^{-123}$, Fisher's exact test with a one-sided alternative hypothesis. **d**, Overlap of class I *daf-16* target genes (see Methods) with genes downregulated in day 10 *spr-4;spr-3;daf-2* triple mutants relative to *daf-2* single mutants. *P* values calculated using a hypergeometric distribution (see Methods). n.s., $P = 0.99$. **e**, Ivermectin increases DAF-16::GFP levels in *spr-4;spr-3* worms following *daf-2* RNAi. Left, confocal imaging of GFP fluorescence in ivermectin-treated (10 pg ml$^{-1}$) and untreated worms. The red dashed lines indicate the worm body. Right, quantification of DAF-16::GFP (mean GFP intensity ± s.e.m., WT/Untreated, $n = 12$; WT/Ivermectin, $n = 10$; *spr-4;spr-3*/Untreated, $n = 10$; *spr-4;spr-3*/Ivermectin, $n = 10$. $^{**}P = 4.6 \times 10^{-4}$ (*spr-4;spr-3* versus WT/untreated), $P = 2.6 \times 10^{-4}$ (*spr-4;spr-3*/Ivermectin versus *spr-4;spr-3*/untreated) by Mann–Whitney *U*-test with multiple testing correction by Holm's method. Shown is a representative experiment replicated three times.



**Extended Data Fig. 10 | Coregulation of FOXO1 and REST in the ageing brain and modulation by glutamatergic signalling. a**, Linear regression analysis of *REST* and *FOXO* mRNA levels in the prefrontal cortex of 174 cognitively intact individuals (age ≥60 years) from the CMC cohort determined by RNA-seq. *P* values derived from a *t*-test for the slope of the regression line and Bonferroni-corrected across all expressed genes. **b**, Colocalization of REST and FOXO1 in neurons of the aged human prefrontal cortex. Confocal immunofluorescence microscopy was performed in human prefrontal cortex using antibodies against REST (green, rabbit polyclonal; Bethyl), FOXO1 (red, goat polyclonal; LS-Bio) and the neuronal marker MAP2 (grey, chicken polyclonal; Abcam). Scale bar, 40 μm. The image shown is representative of immunofluorescence labelling performed in 30 individuals. See Supplementary Table 20 for additional information on antibodies. **c**, Inhibition of glutamatergic signalling in mouse cortical neuronal cultures elevates FOXO1 levels. Left, primary mouse cortical neuronal cultures treated with kynurenic acid (KYNA, 5 μM), AP5 (50 μM), NBQX (2 μM) or vehicle control were analysed by confocal immunofluorescence for FOXO1 or MAP2 and labelled with DAPI. Boxed areas are magnified in the lower three rows. Note that most FOXO1 in cultured neurons is cytoplasmic, but a detectable nuclear component overlaps with DAPI. Scale bar, 40 μm. Right, quantification (mean ± s.e.m.) of total and nuclear FOXO1 levels in MAP2-positive neurons. Control, $n = 200$; KYNA, $n = 326$; AP5, $n = 148$; NBQX, $n = 197$. FOXO1 total/KYNA, $^{**}P = 2.1 \times 10^{-8}$; FOXO1 nuclear/KYNA, $^{**}P = 1.1 \times 10^{-4}$; FOXO1 total/NBQX, $^{**}P = 8.8 \times 10^{-13}$; FOXO1 nuclear/NBQX, $^{**}P = 5.2 \times 10^{-6}$; Mann–Whitney *U*-test with multiple testing correction by Holm's method. Shown is a representative experiment replicated three times.

https://doi.org/10.1038/s41586-019-1652-y

# Decoding human fetal liver haematopoiesis

Dorin-Mirel Popescu[1,21], Rachel A. Botting[1,21], Emily Stephenson[1,21], Kile Green[1], Simone Webb[1], Laura Jardine[1], Emily F. Calderbank[2], Krzysztof Polanski[3], Issac Goh[1], Mirjana Efremova[3], Meghan Acres[1], Daniel Maunder[1], Peter Vegh[1], Yorick Gitton[4], Jong-Eun Park[3], Roser Vento-Tormo[3], Zhichao Miao[3,5], David Dixon[1], Rachel Rowell[1], David McDonald[1], James Fletcher[1], Elizabeth Poyner[1,6], Gary Reynolds[1], Michael Mather[1], Corina Moldovan[7], Lira Mamanova[3], Frankie Greig[1], Matthew D. Young[3], Kerstin B. Meyer[3], Steven Lisgo[8], Jaume Bacardit[9], Andrew Fuller[1], Ben Millar[1], Barbara Innes[1], Susan Lindsay[8], Michael J. T. Stubbington[3], Monika S. Kowalczyk[10], Bo Li[10,11], Orr Ashenberg[10], Marcin Tabaka[10], Danielle Dionne[10], Timothy L. Tickle[10,12], Michal Slyper[10], Orit Rozenblatt-Rosen[10], Andrew Filby[1], Peter Carey[13], Alexandra-Chloé Villani[11,14], Anindita Roy[15], Aviv Regev[10,16], Alain Chédotal[4], Irene Roberts[15,17,18], Berthold Göttgens[2], Sam Behjati[3,19]*, Elisa Laurenti[2]*, Sarah A. Teichmann[3,20]* & Muzlifah Haniffa[1,3,6]*

**Definitive haematopoiesis in the fetal liver supports self-renewal and differentiation of haematopoietic stem cells and multipotent progenitors (HSC/MPPs) but remains poorly defined in humans. Here, using single-cell transcriptome profiling of approximately 140,000 liver and 74,000 skin, kidney and yolk sac cells, we identify the repertoire of human blood and immune cells during development. We infer differentiation trajectories from HSC/MPPs and evaluate the influence of the tissue microenvironment on blood and immune cell development. We reveal physiological erythropoiesis in fetal skin and the presence of mast cells, natural killer and innate lymphoid cell precursors in the yolk sac. We demonstrate a shift in the haemopoietic composition of fetal liver during gestation away from being predominantly erythroid, accompanied by a parallel change in differentiation potential of HSC/MPPs, which we functionally validate. Our integrated map of fetal liver haematopoiesis provides a blueprint for the study of paediatric blood and immune disorders, and a reference for harnessing the therapeutic potential of HSC/MPPs.**

The blood and immune systems develop during early embryogenesis. Our understanding of this process derives from mouse and in vitro model systems, as human fetal tissue is scarce. Although haematopoietic development is conserved across vertebrates[1], there are notable differences between mouse and human[2,3]. Comprehensive interrogation of human tissue to understand the molecular and cellular landscape of early haematopoiesis has implications beyond life in utero, as it provides a blueprint for understanding immunodeficiencies, childhood leukaemias and anaemias and generates insights into HSC/MPP propagation to inform stem cell technologies.

The earliest blood and immune cells originate outside the embryo, arising from the yolk sac between 2 and 3 weeks after conception. At 3–4 post-conception weeks (PCW), intra-embryonic progenitors from the aorta–gonad–mesonephros (AGM) develop[4]. Yolk sac and AGM progenitors colonize fetal tissues such as the liver, which remains the major organ of haematopoiesis until the middle of the second trimester. Fetal bone marrow is colonized around 11 PCW and becomes the dominant site of haematopoiesis after 20 PCW in humans[4]. Yolk sac-, AGM-, fetal liver- and bone marrow-derived immune cells seed peripheral tissues including non-lymphoid tissues (NLTs), where they undergo specific maturation programs that are both intrinsically determined and extrinsically nurtured by the tissue microenvironment[5,6].

In this study, we use single-cell transcriptomics to map the molecular states of human fetal liver cells between 7 and 17 PCW, when the liver is the predominant site of human fetal haematopoiesis. We integrate results from imaging mass cytometry, flow cytometry and cellular morphology to validate the transcriptome-based cellular profiles. We construct the functional organization of the developing immune network by comparative analysis of immune cells in fetal liver with those in yolk sac, and skin and kidney as representative NLTs.

## Single-cell transcriptome of fetal liver

To investigate blood and immune cell development in the fetal liver, we generated single-cell suspensions from embryonic and fetal livers between 7 and 17 PCW. We used fluorescence-activated cell sorting (FACS) to isolate CD45$^+$ and CD45$^-$ cells using adjoining gates for comprehensive capture (Fig. 1a, Extended Data Fig. 9a) for single-cell RNA-sequencing (scRNA-seq) (both 10x Genomics platform and Smart-seq2) (Fig. 1 and Supplementary Table 1). To enable parallel evaluation of blood and immune cell topography in NLT and the yolk

---

[1]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK. [2]Department of Haematology and Wellcome and MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. [4]Sorbonne Université, INSERM, CNRS, Institut de la Vision, Paris, France. [5]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK. [6]Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [7]Department of Pathology, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [8]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. [9]School of Computing, Newcastle University, Newcastle upon Tyne, UK. [10]Klarman Cell Observatory, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [11]Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Boston, MA, USA. [12]Data Sciences Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [13]Haematology Department, Royal Victoria Infirmary, Newcastle-upon-Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [14]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [15]Department of Paediatrics, University of Oxford, Oxford, UK. [16]Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [17]MRC Molecular Haematology Unit and Department of Paediatrics, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. [18]BRC Blood Theme, NIHR Oxford Biomedical Centre, Oxford, UK. [19]Department of Paediatrics, University of Cambridge, Cambridge, UK. [20]Theory of Condensed Matter Group, Cavendish Laboratory/Department of Physics, University of Cambridge, Cambridge, UK. [21]These authors contributed equally: Dorin-Mirel Popescu, Rachel A. Botting, Emily Stephenson. *e-mail: sb31@sanger.ac.uk; el422@cam.ac.uk; st9@sanger.ac.uk; m.a.haniffa@newcastle.ac.uk

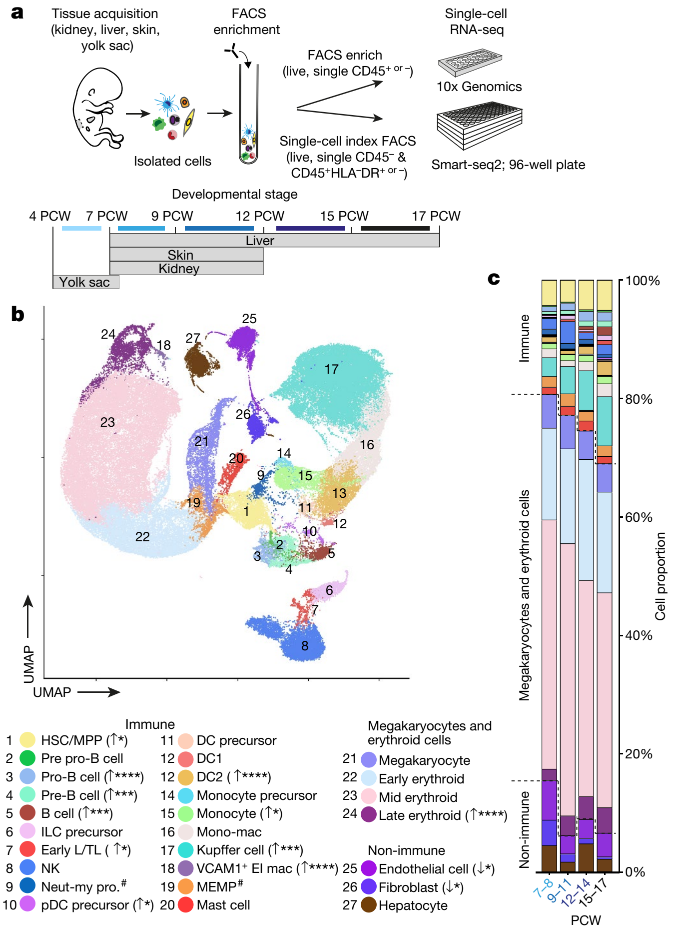


**Fig. 1 | Single-cell transcriptome map of fetal liver. a**, Schematic of tissue processing and cell isolation for scRNA-seq profiling of fetal liver, skin and kidney across four developmental stages (7–8, 9–11, 12–14 and 15–17 PCW), and yolk sac from 4–7 PCW. **b**, UMAP visualization of fetal liver cells from 10x Genomics sequencing using 3′ chemistry. Colours indicate cell state. Mac, macrophage; Neut-my, neutrophil–myeloid; Mono-mac, monocyte-macrophage; Early L/TL, early lymphoid/T lymphocyte; pro., progenitor. Statistical significance of cell frequency change by stage shown in parentheses (negative-binomial regression with bootstrap correction for sort gates; $*P < 0.05$, $***P < 0.001$ and $****P < 0.0001$, details in Supplementary Table 8) with up or down arrows to indicate positive or negative coefficient of change, respectively. #Novel populations. **c**, Liver composition by developmental stage as the mean percentage of each population per stage corrected by CD45$^+$/CD45$^-$ sort fraction. Colours indicate cell states as shown in **b**.

sac during early development (Fig. 1a), we profiled skin, kidney and yolk sac cells using isolation by FACS and the 10x Genomics platform.

In total, 138,575 ($n = 14$) liver (an additional 1,206 cells were profiled using Smart-seq2), 54,690 ($n = 7$) skin, 9,643 kidney ($n = 3$) and 10,071 yolk sac ($n = 3$) cells passed quality control and doublet exclusion (Extended Data Fig. 1a, b, Supplementary Table 2). We performed graph-based Louvain clustering and identified differentially expressed genes (DEGs) to annotate cell clusters. To minimize technical batch effects while preserving biological variation due to gestational stage, we divided liver samples into four gestational-stage categories and performed data integration between samples using Harmony[7] (Extended Data Fig. 1c–e).

We identified 27 major cell states in the fetal liver (Fig. 1b, Extended Data Fig. 1f). VCAM1$^+$ erythroblastic island (EI) macrophages were validated as a distinct cell state owing to their interactions with erythroid cells (Extended Data Fig. 4a–f). We applied a descriptive nomenclature on the basis of gene-expression profiles. All cell states were found throughout the developmental period studied, but the frequency varied according to gestation stage (Fig. 1c, Extended Data Fig. 1e). Neutrophils, basophils and eosinophils were not detected, consistent with reports of granulocytes emerging during fetal bone marrow haematopoiesis[8]. Samples from early stages exhibited erythroid lineage bias, and lymphoid and myeloid lineages were represented at later stages, as previously shown[5] (Fig. 1c, Extended Data Fig. 1g).

Our fetal liver dataset can be explored using an interactive web portal at https://developmentcellatlas.ncl.ac.uk//datasets/hca_liver/. We provide comprehensive expression profiles of genes that are known to cause primary immunodeficiencies[9] to aid future molecular phenotyping of these disorders (Extended Data Fig. 8).

## Validation of genes and cell states

We manually selected 48 genes from the 4,471 DEGs between all clusters (log(fold change) > 0.5) (Fig. 2a). The predictive power of these 48 genes to determine cell states by the Random Forest classifier was 89% on average for precision and recall (Extended Data Fig. 2a). We designed a FACS panel for prospective cell isolation using genes that encode cell surface proteins (Fig. 2a, Extended Data Fig. 9b), which enabled validation of 19 cell types by mini bulk transcriptome profiling using Smart-seq2 (Extended Data Fig. 2b–d); 6 of the 19 cell types were also validated by scRNA-seq. Cytospins from FACS-isolated cells were morphologically consistent with their designated cell type, including hypogranularity of embryonic or fetal mast cells[10] and the resemblance of early erythroid cells to previously reported 'early erythroid progenitors'[11] (Fig. 2b).

Next, we evaluated the spatial distribution of erythroid, mast cell, myeloid and lymphoid lineages using imaging mass cytometry (Extended Data Fig. 2e). The liver architecture evolved considerably between 8 and 15 PCW. Organization of hepatocyte aggregates increased, although hepatic lobules around a central vein and portal triad are not clearly visible. Haematopoietic islands were present in sinusoids and surrounding hepatocyte aggregates. Sinusoidal CD68$^+$ macrophages were surrounded by glycophorin A (GYPA)$^+$ erythroid cells (Extended Data Fig. 2e). CD1c$^+$ dendritic cells (DCs) and CD79a$^+$CD20$^+$ cells from the B cell lineage were sparsely distributed (Extended Data Fig. 2e). The proportions of these cells approximated our scRNA-seq profile for haematopoietic cells but not that for hepatocytes (Extended Data Fig. 2e, Fig. 1b), in keeping with the fragility of hepatocytes following ex vivo isolation and their high expression of mitochondrial genes[12]. By validating our single-cell transcriptome dataset with multiple modalities, we provide an integrated map of haematopoietic cells in the fetal liver.

## Fetal liver and NLT haematopoiesis

Next, we inferred trajectories of haematopoietic development. By force-directed graph (FDG) analysis, we identified three connections to a central HSC/MPP node featuring erythroid–megakaryocyte–mast cells, B cells and innate or T-lymphoid cells and myeloid cells (Fig. 3a and Supplementary Video 1). Partition-based approximate graph abstraction (PAGA) also supported the presence of a shared megakaryocyte–erythroid–mast cell progenitor (MEMP) downstream of HSC/MPPs (Extended Data Fig. 3a). Genes that were dynamically modulated in the specification of erythroid, megakaryocyte and mast cell lineages were distinct: *TAL1* and *KLF1* in the erythroid lineage, *F11R*, *PBX1* and *MEIS1* in the megakaryocyte lineage and *HES1* in mast cell differentiation[13–16] (Extended Data Fig. 3b). We investigated supporting factors for fetal liver erythropoiesis using CellPhoneDB[17] to predict specific or enriched receptor–ligand interactions between erythroblasts and VCAM1$^+$ EI macrophages (Extended Data Fig. 4a). We identified statistically significant interactions for *VCAM1*, *ITGB1*, *ITGA4*, *SIGLEC1*, *ICAM4* and *SPN*, which encode molecules that are known to be important in haematopoiesis[18,19] (Extended Data Fig. 4a). The presence of VCAM1 on EI macrophages and ITGA4 on early-to-mid erythroid cells was confirmed by immunohistochemical analysis on serial fetal liver sections (Extended Data Fig. 4b). Interaction of VCAM1$^+$ EI macrophages with erythroblasts was also observed using

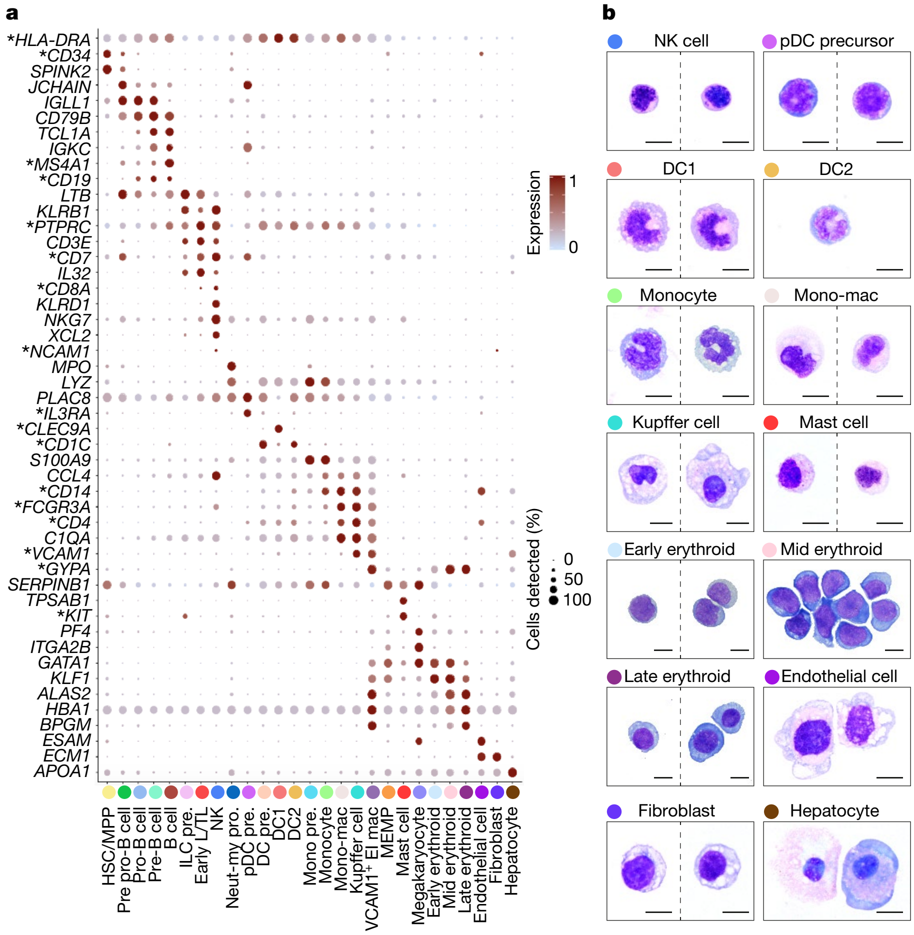


**Fig. 2 | Multi-modal and spatial validation of cell types.** **a**, Median-scaled ln-normalized gene expression of 48 selected DEGs for the liver cell states from Fig. 1b. Asterisks indicate markers used for isolation of cells by FACS. Gene-expression frequency (percentage of cells within each cell type expressing the gene) is indicated by spot size and expression level is indicated by colour intensity. Mono, monocyte; pre., precursor. **b**, Representative Giemsa-stained cytospins showing morphology of populations isolated by FACS on the basis of DEGs marked with asterisks in **a**. Scale bars, 10 μm.

imaging flow cytometry analysis (Extended Data Fig. 4d) and may explain the combined erythroblast and macrophage transcriptome of VCAM1$^+$ EI macrophages (Fig. 2a), which has also been described in mouse central EI macrophages (Extended Data Fig. 4g)[20].

Comparing across haematopoietic tissues, mast cells were also present in the yolk sac (Extended Data Fig. 1b). Erythroblasts show expression of haemoglobin genes and a temporal shift from expression of *HBZ* and *HBE1* to expression of fetal haemoglobin subunits (*HBA1* and *HBG2*) between yolk sac and liver (Fig. 3b).

Megakaryocytes, erythroid cells, mast cells and MEMPs were present in NLTs, but HSC/MPPs were absent (Extended Data Figs. 1a, 3c–e). We compared the highly and differentially expressed genes of corresponding cell types in fetal liver, skin and kidney (Extended Data Fig. 3d). Mast cells, megakaryocytes and cells of the erythroid lineages showed high connectivity (PAGA scores) across all four tissues (Extended Data Fig. 3e). This high connectivity may be a result of local maturation of progenitors in NLT or influx of cells at various differentiation stages. Erythroid cells were absent in kidney, suggesting restricted

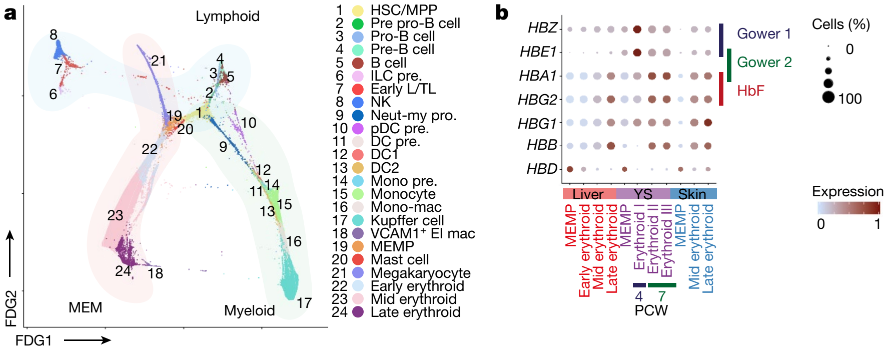


**Fig. 3 | Fetal liver and NLT haematopoiesis.** **a**, FDG visualization of all haematopoietic cells from Fig. 1b. **b**, Dot plot showing the median-scaled ln-normalized expression of globin genes encoding haemoglobin subunits: *HBZ* and *HBE1* (Gower 1), *HBE1* and *HBA1* (Gower 2) and *HBA1* and *HBG2* (fetal; HbF) in liver, skin and yolk sac (YS) erythroid lineages (MEMP and early, mid and late erythroids). Gene-expression frequency (percentage of cells within each cell type expressing the gene) is indicated by spot size and expression level is indicated by colour intensity.

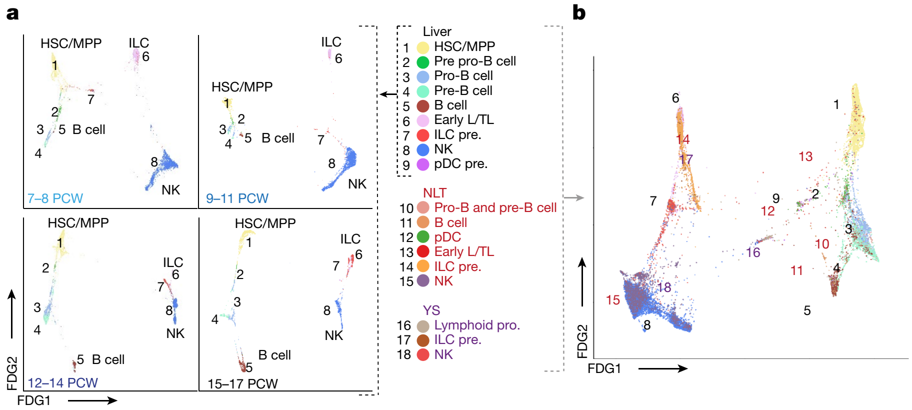


**Fig. 4 | Lymphoid lineages in fetal liver and NLT. a**, FDG visualization of fetal liver HSC/MPP and lymphoid cell types from Fig. 1b showing changes over four developmental stages. **b**, FDG visualization of fetal liver and corresponding skin, kidney and yolk sac lymphoid cells.

differentiation of the MEMP lineage at specific sites (Extended Data Fig. 1a). Immunohistochemical analysis of serial skin sections showed nucleated GYPA$^+$ cells inside and outside CD34$^+$ blood vessels, in keeping with local differentiation of MEMPs (Extended Data Fig. 3f). Light-sheet fluorescence microscopy supported this finding (Extended Data Fig. 3g, Supplementary Video 2). The proliferative capacity of MEMP in NLT was confirmed by expression of *MKI67* and cell cycle genes (Extended Data Fig. 3h). Skin MEMPs expressed some early erythroblast genes including *MYL4* (Extended Data Fig. 3d), suggesting that these may act as erythroid progenitors in the skin[21]. These findings demonstrate that during early development, skin in the physiological state can contribute to erythropoiesis and supplement the erythroid output of fetal liver.

## Lymphoid lineages in fetal liver and NLT

Previous studies have reported the presence of T and B lymphocytes[22], natural killer (NK) cells[23], and innate lymphoid cells (ILCs)[24] in the human fetal liver. In our analysis, we observed two lymphoid branches: an NK–T–ILC lineage and a B lineage (Fig. 4a, Extended Data Fig. 5a). The early lymphoid/T lymphocyte cluster varied by gestational stage, with cells expressing *GATA3*, *KLRB1*, *CD3D*, *CD7* and *JCHAIN* at 7–8 PCW before T cells emerged from the thymus (Fig. 4a, Extended Data Fig. 5b, c). Early in gestation, this cluster may contain fetal liver early thymocyte progenitors, which are capable of generating $\alpha\beta$–T cells when co-cultured with thymic epithelial cells[25,26]. At the 12–14 PCW stage, cells expressed *TRDC* and *TRAC*, but not *GZMB* or *PRF1*, the cytoplasmic granular products characteristic of mature CD8$^+$ T cells. *TRDC* expression was absent at the 15–17 PCW stage (Extended Data Fig. 5b–d). These findings are consistent with seeding of fetal liver by $\gamma\delta$-T cells and $\alpha\beta$–T cells sequentially following their exit from thymus after 12 PCW[27] and are consistent with previous reports of T cell identification only being possible after 18 PCW[26,28].

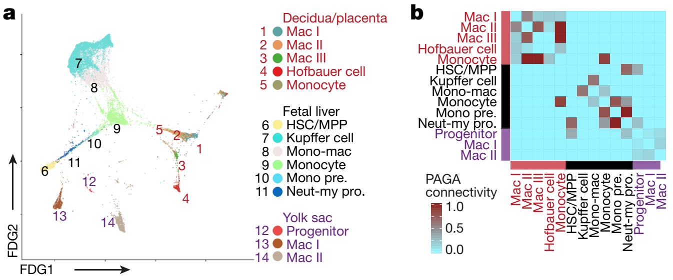


**Fig. 5 | Tissue signatures in developing myeloid cells. a**, FDG visualization of HSC/MPP, myeloid progenitors, monocytes and macrophages from fetal liver, decidua/placenta and yolk sac. **b**, PAGA connectivity scores of the populations shown in **a**.

NK cells (expressing *NCAM1*, *CD7*, *IL2RB* and *CD3E*) and ILC precursors (expressing *KIT*, *KLRB1*, *IL7R* and *RORC*) share a common origin in the lymphoid branch according to PAGA and diffusion-map analyses (Figs. 2a, 4a, Extended Data Fig. 5a, i, Supplementary Table 3). This is consistent with studies showing a shared progenitor of NK and ILCs in human and mouse[29,30].

Cells in the B cell lineage comprised a continuum of differentiation states, from primitive 'pre pro-B' cell clusters expressing *CD34*, *SPINK2* and *IGLL1* to progenitor B (pro-B) and precursor B (pre-B) cell clusters with increasing expression of B cell transcripts *MS4A1*, *CD79B*, *DNTT* and *HLA-DRA* and decreasing expression of *JCHAIN* and *LTB*[31] (Figs. 2a, 4a and Supplementary Table 3). Pro-B and pre-B cell clusters had high nuclear–cytoplasmic ratio and immature chromatin and nucleoli (Extended Data Fig. 5e). Differentially regulated genes in the HSC/MPP-to-B cell transition include *SPIB*, *SP100* and *CTSS* (Extended Data Fig. 5e). Pre-B cells were detected between 7 and 8 PCW, but mature B cells were detected only after 9 PCW (Figs. 1b, c, 4a). We observed a decline in expression of *NFKBIA*, an inhibitor of NF-κB, in HSC/MPPs and an increase in *TNFSF13B* (BAFF) expression in Kupffer cells during gestation (Extended Data Fig. 5f). NF-κB and BAFF are known B cell survival and differentiation factors[32]. The cell-intrinsic versus tissue-microenvironment factors that control B cell differentiation in the fetal liver require further investigation.

Pro-B, pre-B and B cells were present but HSC/MPPs and pre pro-B cells were absent from NLT, in comparison to liver and yolk sac (Fig. 4b, Extended Data Fig. 5g). Early lymphoid/T lymphocytes, NK cells and ILCs in NLT shared a transcriptional signature with their liver counterparts; however, tissue-specific expression of chemokine (*XCLI* and *CXCL8*) and cytotoxic granule (*GNLY*) genes suggest that maturation and tissue adaptation of NK cells take place in the skin and kidney (Extended Data Fig. 5h). ILC precursors in NLT lacked the full set of characteristic markers and transcription factors of their mature progenies ILC1, ILC2 and ILC3 (Extended Data Fig. 5i). NK cells and ILC precursors were present in the yolk sac (Fig. 4b). Together, these findings suggest that NLTs are seeded by NK and ILC precursors from fetal liver and potentially yolk sac, which differentiate in situ and acquire tissue-related gene-expression profiles.

## Tissue signatures in myeloid cells

Fate-mapping studies in mice have demonstrated that tissue macrophages are seeded by yolk sac and fetal liver progenitors[33,34], whereas DCs originate from bone marrow-derived HSC/MPPs through a monocyte-independent lineage[35]. We observed myeloid progenitors, monocytes, macrophages and DC1 and DC2 clusters in fetal liver and NLT as early as 7 PCW (Figs. 1b, 5a, Extended Data Figs. 1a, 6a).

Myeloid lineages stemmed from HSC/MPP via three intermediates: a neutrophil–myeloid progenitor expressing *CD34*, *SPINK2*, *AZU1*, *PRTN3*, *ELANE*, *MPO* and *LYZ*, monocyte precursors and DC

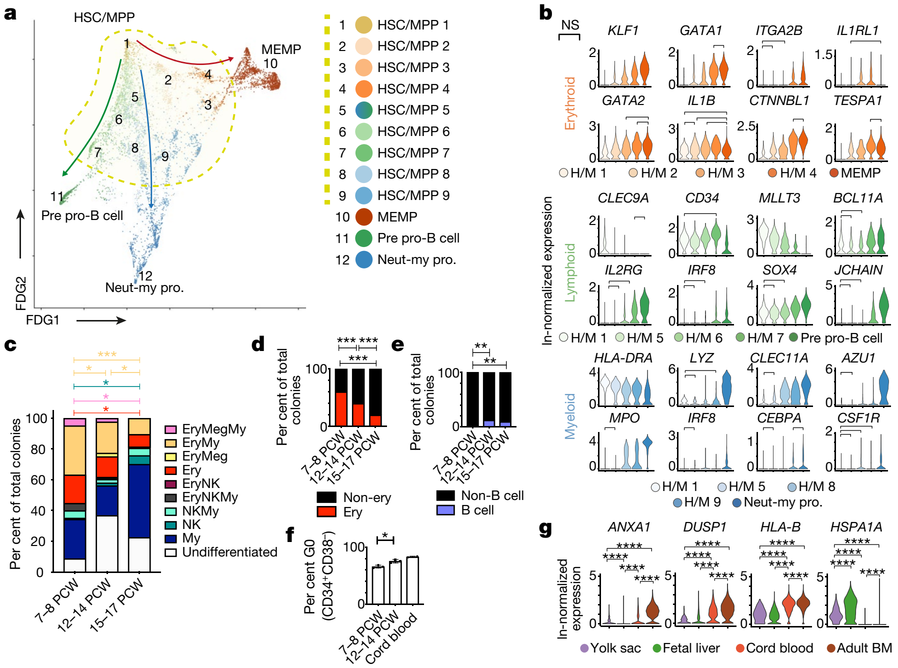


**Fig. 6 | HSC/MPP differentiation potential by gestation.** **a**, FDG visualization of liver HSC/MPP and early haematopoietic progenitor populations from Fig. 1b. **b**, Violin plots showing ln-normalized median gene expression of statistically significant, dynamically variable genes that are up- or downregulated during HSC/MPP transition to neutrophil-myeloid progenitors, MEMP and pre pro-B cells from fetal liver. NS, not significant. H/M, HSC/MPP. **c**, Stacked bar plot of all different types of colonies generated by single 'HSC pool' gated cells in an assay with MS5 stromal cells. $*P < 0.05$, $***P < 0.001$; individual samples shown in Extended Data Fig. 7c. The colour of the bar corresponds to the type of colony tested compared with all others. My, myeloid; ery, erythroid, meg, megakaryocyte. **d**, Percentage of colonies generated by single HSC pool gated cells containing erythroid cells (defined as the sum of ery, ery–meg, ery–meg–my, ery–my, ery–NK and ery–NK–my colonies shown in **c**); $***P < 0.001$. **e**, Percentage of colonies containing B cells following culture in conditions optimized for B and NK cells from 10 cells from the HSC pool gate. $**P < 0.01$. **f**, Percentage of cells in the G0 phase of the cell cycle assessed using Ki67 and DAPI flow cytometry analysis ($*P = 0.0136$). Data are mean ± s.d. **g**, ln-normalized median expression of selected genes in yolk sac progenitors, cord blood HSCs and adult bone marrow (BM) HSCs with significant differential expression compared to fetal liver HSC/MPP, visualized by violin plots ($****P < 0.0001$).

precursors (Figs. 2a, 3a, Supplementary Table 3). DC1 differentiated from neutrophil–myeloid progenitors and DC precursors linked closely to DC2. DC differentiation involved dynamic regulation of *CLEC11A*, *BATF3* and *ID2*, whereas monocyte differentiation involved dynamic regulation of *S100A8/A9*, *FCGR1A/2A* and *S100A12* (Extended Data Fig. 6b). Plasmacytoid DC (pDC) precursors branched from both early myeloid precursors and pre pro-B cells, consistent with recent reports of their mixed lymphoid and myeloid origin in mice[36,37] (Extended Data Fig. 6a). We compared monocytes, macrophages and their putative precursors in placenta, yolk sac and fetal liver (Fig. 5a, b, Extended Data Fig. 6c). The macrophage transcriptome profile was highly tissue-specific and exhibited connectivity to other macrophage subtypes within the tissue of residence, obscuring potential ontogenic relationships (Fig. 5a, b, Extended Data Fig. 6c–e).

Monocytes, macrophages, pDCs, DC1s and DC2s were present in both skin and kidney (Extended Data Fig. 1a). NLT monocytes and DCs correlated strongly with fetal liver counterparts but macrophages were more tissue-specific, with *VCAM1* expression in liver and *F13A1* expression in skin, as observed in adult tissue macrophages[38,39] (Extended Data Fig. 6e). Tissue-specific gene-expression patterns that confer functional specialization were detected for DCs; for example, *S100A4* (encoding a molecule involved in T cell activation) was expressed in skin DC1s[40] and *AOAH* (encoding an enzyme involved in lipopolysaccharide response modulation) was expressed in liver DC2s[41] (Extended Data Fig. 6e). DC activation in spite of the sterile fetal environment suggests an active role for fetal DCs in mediating tolerance, as previously reported[42].

## HSC/MPP differentiation potential by age

Our observation of a HSC/MPP cell state from which the earliest lineage-committed cells radiate is in line with recent observations from scRNA-seq analysis in postnatal mice and humans[36,43,44] (Fig. 6a). At the base of this cluster by FDG visualization was a population expressing *CLEC9A*, *HLA-DRA* and the highest levels of primitive genes including *MLLT3*, consistent with a multipotent long-term repopulating (LT)-HSC population[45] (Fig. 6a, b). Using DEGs and supervised analysis, HSC/MPP clusters with gene expression intermediate between LT-HSCs and early progenitors similar to human lymphoid-primed short-term (ST)-HSCs[45], the mouse erythroid-biased MMP2 and myeloid-biased MMP3[46,47] were identifiable, demonstrating early transcriptome priming along all differentiation branches within the MPP pool (Fig. 6a, b).

We hypothesized that the cellular composition of the developing fetal liver resulted from local modulation of HSC/MPP potential. To test this, we used FACS to isolate single cells from the $CD34^+CD38^+$, $CD34^+CD38^-CD45RA^-$ and $CD34^+CD38^-CD45RA^+$ FACS gates and profiled them by plate-based single-cell transcriptomics (Smart-seq2) and single-cell clonal differentiation assays[45,48] (Extended Data Fig. 7a). A support vector machine trained on the fetal liver dataset identified enrichment of HSC/MPPs (to about 85%) in the $CD34^+CD38^-$ gate, with the majority of cells in the $CD34^+CD38^-CD45RA^+$ multilymphoid progenitor (MLP) gate also classified as HSC/MPP (Extended Data Fig. 7b). This is in agreement with reported transcriptional similarity between HSC/MPP and MLP[49] and our identification of lymphoid priming in the HSC/MPP compartment. Single-cell culture from $CD34^+CD38^-CD45RA^-$ HSC/MPPs yielded both uni- and multipotential colonies (Fig. 6c–e, Extended Data Figs. 7c–g, 9c, d). There was a significant reduction in trilineage colonies with gestational age (Extended Data Fig. 7f). Colonies containing erythroid cells significantly decreased in number, whereas those containing NK cells and B cells increased with gestational age (Fig. 6c–e, Extended Data Fig. 7c–g). HSC/MPP from embryonic livers at less than 9 PCW generated almost no B cells (Fig. 6e), in keeping with the paucity of B cells at this stage (Figs. 1c, 4a, Extended Data Fig. 1g). These findings support the hypothesis of differential HSC/MPP intrinsic potential according to gestational stage and mirror our observation of early erythroid predominance and greater lymphoid representation at later stages (Figs. 1c, 4a).

Comparing HSC/MPPs and early progenitors across haematopoietic tissues during development, higher *MKI67* expression and cell cycle staining suggest enhanced proliferative potential of yolk sac and fetal liver progenitors and HSC/MPPs[50–52] (Extended Data Fig. 7h). The fraction of fetal liver HSC/MPPs in G0 increased with gestational age (Fig. 6f, Extended Data Fig. 7i), indicating a progressive shift to quiescence during fetal life. Fetal liver HSC/MPPs showed higher expression of genes encoding a heat shock protein (*HSPA1A*), potentially for maintenance of genome and proteome integrity, and lower levels of MHC-I (*HLA-B*), suggesting reduced antigen-presenting potential compared with cord blood and adult bone marrow HSC/MPPs (Fig. 6g).

Collectively, our findings demonstrate that intrinsic changes in HSC/MPP numbers, proliferation and differentiation potential occur over the first and second developmental trimesters. These changes are likely to be pivotal for fetal liver haematopoiesis to adapt to the needs of the developing fetus; first the establishment of an effective oxygen transport system and subsequently the development of a complete blood and immune system.

## Discussion

Development of the human immune system in utero has remained poorly understood. Using single-cell transcriptome profiling, we have resolved the cellular heterogeneity and abstract dynamic temporal information on blood and immune development in fetal liver, yolk sac, skin and kidney. Large-scale scRNA-seq studies during human development must account for technical batch effects without compromising detection of biological variations over gestation. Our approach highlights key insights: physiological erythropoiesis in fetal skin, establishment of a DC network as early as 7 PCW, potential dual myeloid and lymphoid origin of pDCs, seeding of mast cells, NK cells and ILCs from the yolk sac and tissue adaptation of NK cells, ILCs, DCs and macrophages during development. Our findings reveal modulation of HSC/MPP intrinsic differentiation potential over gestational age, suggesting that this may be an additional functional mechanism to regulate haematopoietic output of the fetal liver throughout the first and second trimesters.

In summary, our comprehensive fetal liver atlas provides a foundational resource for understanding fetal liver haematopoiesis and the developing immune system. Our reference dataset will be invaluable for studies on paediatric blood and immune disorders and exploiting HSC/MPPs for therapy. Our approach using single-cell transcriptomics to study human development provides a framework that can be applied to study any temporal processes across the human lifespan.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1652-y.

Received: 26 October 2018; Accepted: 9 September 2019;
Published online 9 October 2019.

1. Jagannathan-Bogdan, M. & Zon, L. I. Hematopoiesis. *Development* **140**, 2463–2467 (2013).
2. Parekh, C. & Crooks, G. M. Critical differences in hematopoiesis and lymphoid development between humans and mice. *J. Clin. Immunol.* **33**, 711–715 (2013).
3. Ivanovs, A. et al. Human haematopoietic stem cell development: from the embryo to the dish. *Development* **144**, 2323–2337 (2017).
4. Holt, P. G. & Jones, C. A. The development of the immune system during pregnancy and early life. *Allergy* **55**, 688–697 (2000).
5. Kashem, S. W., Haniffa, M. & Kaplan, D. H. Antigen-presenting cells in the skin. *Annu. Rev. Immunol.* **35**, 469–499 (2017).
6. Mass, E. et al. Specification of tissue-resident macrophages during organogenesis. *Science* **353**, aaf4238 (2016).
7. Korsunsky, I. et al. Fast, sensitive, and accurate integration of single cell data with Harmony. Preprint at *bioRxiv* https://doi.org/10.1101/461954 (2018).
8. Ohls, R. K. et al. Neutrophil pool sizes and granulocyte colony-stimulating factor production in human mid-trimester fetuses. *Pediatr. Res.* **37**, 806–811 (1995).
9. Picard, C. et al. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. *J. Clin. Immunol.* **38**, 96–128 (2018).
10. Gentek, R. et al. Hemogenic endothelial fate mapping reveals dual developmental origin of mast cells. *Immunity* **48**, 1160–1171 (2018).
11. Iskander, D. et al. Elucidation of the EP defect in Diamond–Blackfan anemia by characterization and prospective isolation of human EPs. *Blood* **125**, 2553–2557 (2015).
12. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
13. An, X. et al. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* **123**, 3466–3477 (2014).
14. Gautier, E.-F. et al. Comprehensive proteomic analysis of human erythropoiesis. *Cell Rep.* **16**, 1470–1484 (2016).
15. Dedhia, P., Kambayashi, T. & Pear, W. S. Notch2 paves the way to mast cells by Hes1 and Gata3. *Proc. Natl Acad. Sci. USA* **105**, 7629–7630 (2008).
16. Okada, Y. et al. Homeodomain proteins MEIS1 and PBXs regulate the lineage-specific transcription of the platelet factor 4 gene. *Blood* **101**, 4748–4756 (2003).
17. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
18. Klei, T. R. L., Meinderts, S. M., van den Berg, T. K. & van Bruggen, R. From the cradle to the grave: the role of macrophages in erythropoiesis and erythrophagocytosis. *Front. Immunol.* **8**, 73 (2017).
19. Kessel, K. U. et al. Emergence of CD43-expressing hematopoietic progenitors from human induced pluripotent stem cells. *Transfus. Med. Hemother.* **44**, 143–150 (2017).
20. Li, W. et al. Identification and transcriptome analysis of erythroblastic island macrophages. *Blood* **134**, 480–491 (2019).
21. Ebert, B. L. et al. An erythroid differentiation signature predicts response to lenalidomide in myelodysplastic syndrome. *PLoS Med.* **5**, e35 (2008).
22. Gale, R. P. in *Fetal Liver Transplantation* (eds Touraine, J.-L., Gale, R. P. & Kochupillai, V.) 45–56 (Springer, 1987).
23. Phillips, J. H. et al. Ontogeny of human natural killer (NK) cells: fetal NK cells mediate cytolytic function and express cytoplasmic CD3ε,δ proteins. *J. Exp. Med.* **175**, 1055–1066 (1992).
24. Forkel, M. et al. Composition and functionality of the intrahepatic innate lymphoid cell-compartment in human nonfibrotic and fibrotic livers. *Eur. J. Immunol.* **47**, 1280–1294 (2017).
25. Haynes, B. F. & Heinly, C. S. Early human T cell development: analysis of the human thymus at the time of initial entry of hematopoietic stem cells into the fetal thymic microenvironment. *J. Exp. Med.* **181**, 1445–1458 (1995).
26. Sánchez, M. J., Spits, H., Lanier, L. L. & Phillips, J. H. Human natural killer cell committed thymocytes and their relation to the T cell lineage. *J. Exp. Med.* **178**, 1857–1866 (1993).
27. Darrasse-Jèze, G., Marodon, G., Salomon, B. L., Catala, M. & Klatzmann, D. Ontogeny of CD4$^+$CD25$^+$ regulatory/suppressor T cells in human fetuses. *Blood* **105**, 4715–4721 (2005).
28. Wucherpfennig, K. W. et al. Structural requirements for binding of an immunodominant myelin basic protein peptide to DR2 isotypes and for its recognition by human T cell clones. *J. Exp. Med.* **179**, 279–290 (1994).
29. Spits, H. et al. Innate lymphoid cells—a proposal for uniform nomenclature. *Nat. Rev. Immunol.* **13**, 145–149 (2013).
30. Chen, L. et al. CD56 expression marks human group 2 innate lymphoid cell divergence from a shared NK cell and group 3 innate lymphoid cell developmental pathway. *Immunity* **49**, 464–476 (2018).
31. Roy, A. et al. Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21. *Proc. Natl Acad. Sci. USA* **109**, 17579–17584 (2012).
32. Almaden, J. V. et al. B-cell survival and development controlled by the coordination of NF-κB family members RelB and cRel. *Blood* **127**, 1276–1286 (2016).
33. Stremmel, C. et al. Yolk sac macrophage progenitors traffic to the embryo during defined stages of development. *Nat. Commun.* **9**, 75 (2018).
34. Ginhoux, F. & Jung, S. Monocytes and macrophages: developmental pathways and tissue homeostasis. *Nat. Rev. Immunol.* **14**, 392–404 (2014).
35. Murphy, T. L. et al. Transcriptional control of dendritic cell development. *Annu. Rev. Immunol.* **34**, 93–119 (2016).
36. Tusi, B. K. et al. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* **555**, 54–60 (2018).
37. Rodrigues, P. F. et al. Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. *Nat. Immunol.* **19**, 711–722 (2018).
38. Seu, K. G. et al. Unraveling macrophage heterogeneity in erythroblastic islands. *Front. Immunol.* **8**, 1140 (2017).
39. McGovern, N. et al. Human dermal CD14$^+$ cells are a transient population of monocyte-derived macrophages. *Immunity* **41**, 465–477 (2014).
40. Sun, J.-B. et al. Deficiency in calcium-binding protein S100A4 impairs the adjuvant action of cholera toxin. *Front. Immunol.* **8**, 1119 (2017).
41. Janelsins, B. M., Lu, M. & Datta, S. K. Altered inactivation of commensal LPS due to acyloxyacyl hydrolase deficiency in colonic dendritic cells impairs mucosal Th17 immunity. *Proc. Natl Acad. Sci. USA* **111**, 373–378 (2014).

42. McGovern, N. et al. Human fetal dendritic cells promote prenatal T-cell immune suppression through arginase-2. *Nature* **546**, 662–666 (2017).
43. Grün, D. et al. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016).
44. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
45. Belluschi, S. et al. Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat. Commun.* **9**, 4100 (2018).
46. Cabezas-Wallscheid, N. et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522 (2014).
47. Pietras, E. M. et al. Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell* **17**, 35–46 (2015).
48. Doulatov, S. et al. Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors. *Cell Stem Cell* **13**, 459–470 (2013).
49. Laurenti, E. et al. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756 (2013).
50. Morrison, S. J., Hemmati, H. D., Wandycz, A. M. & Weissman, I. L. The purification and characterization of fetal liver hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **92**, 10302–10306 (1995).
51. Bowie, M. B. et al. Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *J. Clin. Invest.* **116**, 2808–2816 (2006).
52. Copley, M. R. et al. The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. *Nat. Cell Biol.* **15**, 916–925 (2013).

## METHODS

The experiments were not randomized. Unless otherwise stated, the investigators were not blinded to allocation during experiments and outcome assessment.

**Tissue acquisition.** Human developmental tissues were obtained from the MRC–Wellcome Trust-funded Human Developmental Biology Resource (HDBR; http://www.hdbr.org)[53] with appropriate written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee (08/H0906/21+5). HDBR is regulated by the UK Human Tissue Authority (HTA; www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice.

Embryos and fetal specimens used for light-sheet fluorescence microscopy were obtained with written informed consent from the parents (Gynaecology Hospital Jeanne de Flandres, Lille, France) with the approval of the local ethics committee (protocol no. PFS16-002). Tissues were made available in accordance with the French bylaw (Good practice concerning the conservation, transformation and transportation of human tissue to be used therapeutically, published on December 29, 1998). Permission to use human tissues was obtained from the French agency for biomedical research (Agence de la Biomédecine, Saint-Denis La Plaine, France). No statistical methods were used to predetermine sample size.

**Tissue processing.** All tissues were processed immediately after isolation using the same protocol. Tissue was transferred to a sterile 10 mm$^2$ tissue culture dish and cut into $<$1 mm$^3$ segments before being transferred to a 50-ml conical tube. Yolk sac content was aspirated for analysis before yolk sac digestion. Tissue was digested with 1.6 mg ml$^{-1}$ collagenase type IV (Worthington) in RPMI (Sigma-Aldrich) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Gibco), 100 U ml$^{-1}$ penicillin (Sigma-Aldrich), 0.1 mg ml$^{-1}$ streptomycin (Sigma-Aldrich), and 2 mM L-glutamine (Sigma-Aldrich) for 30 min at 37 °C with intermittent shaking. Digested tissue was passed through a 100-μm filter, and cells were collected by centrifugation (500*g* for 5 min at 4 °C). Cells were treated with 1× RBC lysis buffer (eBioscience) for 5 min at room temperature and washed once with flow buffer (PBS containing 5% (v/v) FBS and 2 mM EDTA) before counting.

**Fetal developmental stage assignment and chromosomal assessment.** Embryos up to 8 PCW were staged using the Carnegie staging method[54]. After 8 PCW, developmental age was estimated from measurements of foot length and heel-to-knee length and compared against a standard growth chart[55]. A piece of skin, or where this was not possible, chorionic villi tissue was collected from every sample for quantitative PCR analysis using markers for the sex chromosomes and autosomes 13, 15, 16, 18, 21, 22, which are the most commonly seen chromosomal abnormalities. All samples were karyotypically normal.

**Flow cytometry and FACS for scRNA-seq.** Antibody panels were designed to allow enrichment of cell fractions for sequencing and cell type validation. Antibodies used for FACS isolation are listed in Supplementary Table 14. An antibody cocktail was prepared fresh by adding 3 μl of each antibody in 50 μl Brilliant Stain Buffer (BD Biosciences) per tissue. Cells ($<10 \times 10^6$) were resuspended in 50–100 μl flow buffer and an equal volume of antibody mix was added to cells from each tissue. Cells were stained for 30 min on ice, washed with flow buffer and resuspended at $10 \times 10^6$ cells per ml. DAPI (Sigma-Aldrich) was added to a final concentration of 3 μM immediately before sorting. Flow sorting was performed on a BD FACSAria Fusion instrument using DIVA v.8, and data were analysed using FlowJo (v.10.4.1, BD Biosciences). Cells were gated to exclude dead cells and doublets, and then isolated for scRNA-seq analysis (10x or Smart-seq2). For 10x sequencing, cells were sorted into chilled FACS tubes coated with FBS and prefilled with 500 μl sterile PBS. For Smart-seq2, single cells were index-sorted into 96-well LoBind plates (Eppendorf) containing 10 μl lysis buffer (TCL (Qiagen) + 1% (v/v) β-mercaptoethanol) per well. B cells were also investigated by flow cytometry as per a previous study[56].

**Cytospins and mini bulk RNA-seq validation.** Fetal liver cells were immunostained with two separate panels (see Supplementary Table 15 for antibody details). Cells were stained for 30 min on ice followed by DAPI staining. FACS was performed on a BD FACSAria Fusion instrument, and data were analysed using FlowJo (v.10.4.1). Cells were isolated into chilled FACS tubes coated with FBS and prefilled with 500 μl sterile PBS for cytospin (500–2,000 cells), or into 1.5-ml microfuge tubes containing 20 μl lysis buffer (100 cells). Giemsa staining (Sigma-Aldrich) was used to determine the morphology of sorted cells on cytospins. Slides were viewed using a Zeiss AxioImager microscope, images taken of 4 fields from $n = 3$ samples using the 100× objective, and viewed using Zen (v.2.3). Raw images are supplied in the Supplementary Information.

**HSC/MPP culture.** MS5 in log-phase growth (DSMZ, passage 6–10, or imported from K. Itoh at Kyoto University; all tested mycoplasma-free, not authenticated) were seeded into 96-well flat-bottom plates (Nunclon delta surface; Thermo) at a density of 3,000 cells per well 24 h before sorting. Medium was Myelocult H5100 (Stem Cell Technologies) supplemented with 100 U ml$^{-1}$ penicillin and 0.1 mg ml$^{-1}$ streptomycin (Sigma-Aldrich). On the day of sorting, medium was replaced with Stem Pro-34 SFM medium (Life Technologies) supplemented with 100 U ml$^{-1}$ penicillin and 0.1mg ml$^{-1}$ streptomycin, 2 mM L-glutamine (Sigma-Aldrich), 100 ng ml$^{-1}$ stem cell factor (Miltenyi), 20 ng ml$^{-1}$ Flt3 (Miltenyi), 100 ng ml$^{-1}$ TPO (Miltenyi), 3 ng ml$^{-1}$ EPO (Eprex), 50 ng ml$^{-1}$ IL6 (Miltenyi), 10 ng ml$^{-1}$ IL3 (Miltenyi), 50 ng ml$^{-1}$ IL11 (Miltenyi), 20 ng ml$^{-1}$ GM-CSF (Miltenyi), 10 ng ml$^{-1}$ IL2 (Miltenyi), IL7 20 ng ml$^{-1}$ (Miltenyi) and 50 ng ml$^{-1}$ lipids (hLDL) (Life Technologies) as described previously[45].

Frozen fetal liver cells were thawed and stained with 10 μl per 10$^6$ cells of antibody cocktail (see Supplementary Tables 16, 17 for antibody details) for 30 min on ice. Three populations of HSC/MPPs and progenitors were isolated from fetal liver suspension. Populations were identified from the DAPI$^-$, doublet-excluded gate as CD3$^-$CD16$^-$CD11c$^-$CD14$^-$CD19$^-$CD56$^-$CD34$^+$ cells (see Supplementary Table 16 for antibody details). The HSC/MPP pool and MLP were found within the 20% of cells with lowest CD38 expression: the HSC/MPP pool were CD90$^{+/-}$ and CD45RA$^-$, whereas MLP were CD90$^-$CD45RA$^+$. Progenitors with the highest 20% of CD38 expression were sorted for comparison. Single cells were sorted using a BD FACSAria Fusion, and sorted directly onto MS5 or medium for culture, or into 96-well LoBind plates containing 10 μl per well lysis buffer for Smart-seq2 scRNaseq (Supplementary Table 12). Analysis of single-cell-derived colonies was performed as described[45]. In brief, colonies were collected into 96 U-bottom plates using a plate filter to prevent carryover of MS5 cells. Cells were stained with 50 μl per well of antibody cocktail (Supplementary Table 16, 17 for antibody details), incubated for 20 min in the dark at room temperature and then washed with 100 μl per well of PBS + 3% FBS. The type (lineage composition) and the size of the colonies formed were assessed by high-throughput flow cytometry (BD FACS Symphony). Colony output was determined using the gating strategy shown in Extended Data Fig. 9c. A single cell was defined as giving rise to a colony if the sum of cells detected in the CD45$^+$ and GYPA$^+$ gates was $\geq$ 30 cells. Erythroid colonies were identified as CD45$^-$GYPA$^+$ $\geq$ 30 cells, megakaryocyte colonies as CD41$^+$ $\geq$ 30 cells, myeloid colonies as $[(\text{CD45}^+\text{CD14}^+) + (\text{CD45}^+\text{CD15}^+)] \geq 30$ cells, NK colonies as CD45$^+$CD56$^+$ $\geq$ 30 cells. All high-throughput screening flow cytometry data were recorded in a blinded way, and correlation between the colony phenotype and originating population was only performed at the final stage. Two-tailed Fisher's exact tests, performed in Prism (v.8.1.2, GraphPad Software), were applied to the numbers of colonies of each type by stage to determine statistical significance in lineage differentiation potential with development. For differentiation of B cells from the HSC/MPP pool, ten cells were sorted directly on MS5 stroma in the same conditions as described previously[48] (see Supplementary Table 17 for antibody details). An example of the gating strategy is shown in Extended Data Fig. 9d. For Ki67 staining, cells were stained using the antibody panel in Supplementary Table 18, CD34$^+$ cells were sorted, fixed and permeabilized using the BD Cytofix/Cytoperm kit according to the manufacturer's instructions and then stained overnight for Ki67-FITC followed by DAPI as a DNA dye.

**ImageStream analysis of fetal liver cell suspensions.** Frozen fetal liver cells were thawed and stained with the antibody cocktail (see Supplementary Table 19 for antibody details) for 30 min on ice. Cells were washed with flow buffer and resuspended at the same cell density that was used for cell-sorting experiments ($10 \times 10^6$ cells per ml). DAPI (Sigma-Aldrich) was added to a final concentration of 3 μM immediately before acquisition to identify and exclude dead cells from the experiment. Samples were acquired on a fully calibrated ImageStream X MKII system (Luminex Corporation) using 488 nm, 561 nm, 405 nm and 642 nm excitation lasers and the 60× magnification collection optic. Laser powers were set in order to maximize signal resolution but minimize any saturation of the CCD camera with bright-field images collected in channels 1 and 9. A minimum of 50,000 cell events were collected per sample. In order to calculate spectral compensation, single-stained antibody capture beads were acquired with the bright-field illumination turned off. Spectral compensation and data analysis were performed using the IDEAS analysis software (v.6.2.64, Luminex Corp). In brief, dead cells were first excluded based on DAPI positivity. Overt doublets and debris were excluded from the live cell population using the aspect ratio and area of the bright-field image in channel 1. Single cells had an intermediate area value and an aspect ratio between 1 and 0.6 and were gated as such. Overt doublets present as having an increased area and a lower aspect ratio value and were also gated as such. In all cases, the position and boundaries of a gate were checked for appropriateness using the underlying imagery. The key cell types within the sample were then identified and gated using the total integrated (post-compensation) fluorescence signals from each labelled antibody in the panel in an analogous fashion to conventional flow cytometry data. Each major cell type was then interrogated using the associated multi-spectral images for true single-cell identity or for the presence of 'pernicious doublets' (cells with either significant debris attached or large cells with much smaller cells attached). The same phenotypic analysis was extended to the cells in the overt doublet gate. Further analysis was performed in FlowJo (v.10.4.1). Kruskal–Wallis tests with Dunn's post hoc test, performed in Prism (v.8.1.2, GraphPad Software), were applied to the proportion of each cell type found within the doublet gate to determine statistical significance in doublet rates across fetal liver cell types.

**Library preparation and sequencing.** For the droplet-encapsulation scRNA-seq experiments, 7,000 live, single, CD45$^+$ or CD45$^-$ FACS-isolated cells were loaded onto each channel of a Chromium chip before encapsulation on the Chromium Controller (10x Genomics). Single-cell sequencing libraries were generated using the Single Cell 3′ v.2 and V(D)J Reagent Kits (for T cell receptor repertoire analysis) as per the manufacturer's protocol. Libraries were sequenced using an Illumina HiSeq 4000 using v.4 SBS chemistry to achieve a minimum depth of 50,000 raw reads per cell. The libraries were sequenced using the following parameters: Read1: 26 cycles, i7: 8 cycles, i5: 0 cycles; Read2: 98 cycles to generate 75-bp paired-end reads.

For the plate-based scRNA-seq experiments, a slightly modified Smart-seq2 protocol was used as previously described[57]. After cDNA generation, libraries were prepared (384 cells per library) using the Illumina Nextera XT kit (Illumina). Index v.2 sets A, B, C and D were used per library to barcode each cell before multiplexing. Each library was sequenced to achieve a minimum depth of one-to-two-million raw reads per cell using an Illumina HiSeq 4000 using v.4 SBS chemistry to generate 75-bp paired-end reads.

For the mini bulk RNA-seq experiments, each cell lysate was transferred into a 96-well lo-bind plate (Eppendorf) then processed using the same modified Smart-seq2 protocol as described above. After cDNA generation, libraries were prepared using the Illumina NexteraXT kit with Index v.2 set A to barcode each mini bulk library before multiplexing. All libraries were sequenced on one lane of an Illumina HiSeq 4000 using v.4 SBS chemistry to generate 75-bp paired-end reads and aiming to achieve a minimum depth of 10 million reads per library.

**Immunohistochemistry.** Formalin-fixed, paraffin-embedded blocks of fetal livers aged 6 PCW, 8 PCW, 10 PCW and 13 PCW were obtained from the HDBR. Each was sectioned at 4-μm thickness onto APES-coated slides. Sections were dewaxed for 5 min in Xylene (Fisher Chemical) then rehydrated through graded ethanol (99%, 95% and 70%; Fisher Chemical) and washed in running water. Sections were treated with hydrogen peroxide block (1% v/v in water; Sigma) for 10 min and rinsed in water before antigen retrieval. Citrate antigen retrieval was used for all sections. Citrate buffer, pH 6 was used with pressure heating for antigen retrieval, and then slides placed in TBS, pH 7.6 for 5 min before staining. Staining was done using the Vector Immpress Kit (Vector Laboratories). Sections were blot dried and blocked sequentially with 2.5% normal horse serum, avidin (Vector Laboratories) and then biotin (Vector Laboratories) for 10 min each and blot dried in between. Sections were incubated for 60 min with primary antibody diluted in TBS pH 7.6 (see Supplementary Table 20 for antibody details). Slides were washed twice in TBS pH 7.6 for 5 min each before incubation for 30 min with the secondary antibody supplied with the kit. Slides were washed twice in TBS pH 7.6 for 5 min each, and developed using peroxidase chromogen DAB. Sections were counterstained in Mayer's haematoxylin for 30 s, washed and put in Scott's tap water for 30 s. Slides were dehydrated through graded ethanol (70% to 99%) and then placed in xylene before mounting with DPX (Sigma-Aldrich). Sections were imaged on a Nikon Eclipse 80i microscope using NIS-Elements Fv4.

**Alignment, quantification and quality control of scRNA-seq data.** Droplet-based (10x) sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite (v.2.0.2, 10x Genomics) using the GRCh38 human reference genome (official Cell Ranger reference, v.1.2.0). Smart-seq2 sequencing data were aligned with STAR (v.2.5.1b), using the STAR index and annotation from the same reference as the 10x data. Gene-specific read counts were calculated using htseq-count (v.0.10.0). Cells with fewer than 200 detected genes and for which the total mitochondrial gene expression exceeded 20% were removed. Genes that were expressed in fewer than 3 cells were also removed. We detected on average ~3,000 genes per cell with the 10x Genomics platform and ~6,000 genes with the Smart-seq2 protocol.

**Doublet detection.** Doublets were detected with an approach adapted from a previous study[58]. In the first step of the process, each 10x lane was processed independently. Scrublet[59] was run, obtaining per-cell doublet scores. The standard Seurat-inspired Scanpy processing pipeline was performed up to the clustering stage, using default parameters. Each cluster was subsequently separately clustered again, yielding an over-clustered manifold, and each of the resulting clusters had its Scrublet scores replaced by the median of the observed values. The resulting scores were assessed for statistical significance, with *P* values computed using a right-tailed test from a normal distribution centred on the score median and a median absolute deviation (MAD)-derived standard deviation estimate. The MAD was computed from above-median values to circumvent zero truncation. The *P* values were corrected for false discovery rate with the Benjamini–Hochberg procedure, and a significance threshold of 0.1 was imposed. In the second step of the process, all 10x lanes for a single tissue were pooled together and the Seurat-inspired Scanpy processing was repeated, with the addition of Harmony[7] with a theta of 3 for batch correction between the lanes before the neighbour graph identification step. The joint manifold was clustered, and the frequency of identified doublets was computed. The same statistical framework as in the first step was used to identify clusters significantly enriched in doublets, which were subsequently flagged as doublets in their entirety and removed.

**Clustering and annotation.** Downstream analysis included data normalization (NormalizeData, LogNormalize method, scaling factor 10,000), data feature scaling (ScaleData), variable gene detection (FindVariableGenes), PCA (RunPCA, from variable genes) and Louvain graph-based clustering (FindClusters, data dimensionality reduction using PCA, clustering resolution (res.30)) performed using the R package Seurat (v.2.3.4). Cluster cell identity was assigned by manual annotation using known marker genes and computed DEGs using the FindAllMarkers function in the Seurat package (one-tailed Wilcoxon rank sum test, *P* values adjusted for multiple testing using the Bonferroni correction; Supplementary Table 3). For computing DEGs, all genes were probed provided they were expressed in at least 25% of cells in either of the two populations compared and the expression difference on a natural log scale was at least 0.25. Manual annotation was performed iteratively, which included validating proposed cell labels with known markers and further investigating clusters for which the gene signatures indicated additional diversity. Numbers of each cell type per sample, annotations per cell, and nGene and nUMI per cell type are reported in Supplementary Tables 4, 5, 8 and 9, respectively.

Clustering and cell type assignment for fetal liver data was assessed using two additional clustering methods (not shown): agglomerative clustering (with Ward linkage and Euclidean affinity) and Gaussian mixture (AgglomerativeClustering class from the cluster module and GaussianMixture from the mixture module in sklearn v.0.19.1 Python 3.6.3). Consensus agreement between the three clustering methods was measured by Rand index and adjusted mutual information implemented in the metrics module in the sklearn package. The Rand index scores were 0.89 and 0.85 for agglomerative and Gaussian mixture clustering methods, respectively.

After annotation was completed, a cell type classifier was built by training a support vector machine (SVM) on labelled fetal liver scRNA-seq data with grid search for parameter optimization based on training data. Seventy per cent of the data was used for training and the other 30% for test. The SVM was previously compared in terms of accuracy and recall with a random forest and logistic regression classifiers trained on the same data. Out of the three classifiers, the SVM was chosen because it showed a mean accuracy and weighted mean recall of 95%. Random forest showed 89% for both precision and recall (Supplementary Table 11). The SVM classifier was used for automatic annotation of the Smart-seq2 and mini bulk RNA sequencing datasets to enable identification of biologically meaningful clusters and DEG computation.

Data generated from fetal skin, kidney and yolk sac were pre-processed, normalized, clustered and manually annotated in parallel with, and using the same pipeline as, the liver data. Annotation by cell type for skin and kidney, and yolk sac are reported in Supplementary Tables 6 and 7, respectively. Skin and kidney data were combined using the MergeSeurat function. Clusters characterized by differentially expressed immune gene markers were extracted from the NLT dataset for subsequent comparative analysis with liver-derived immune populations. Human cord blood and adult bone marrow datasets were downloaded from the Human Cell Atlas data portal (https://data.humancellatlas.org/). These were processed using the same approach as described above, followed by manual annotation. Decidua and placental data from ref. [17] were downloaded from ArrayExpress record E-MTAB-6701.

**Data integration.** We used Harmony data integration[7] to correct for batch effect between sample identities. The average kBET rejection rate statistically significantly improved from 0.735 to 0.471 (Supplementary Table 13) following Harmony data integration ($P = 3.83 \times 10^{-3}$ in Kolmogorov–Smirnov test and $P = 8.8 \times 10^{-6}$ in Wilcoxon signed-rank test). The manifold was subjected to re-clustering using Harmony adjusted principal components (PCs) with parameters as mentioned above in 'Clustering and annotation'. Cell-type classifications were then ascertained through re-annotation of the clusters derived from Harmony-adjusted PCs to produce the final annotation.

**Changes in cell proportions over development.** Comparison of cell proportions across gestational stages was assessed by modelling cell number data with negative binomial regression based on Poisson-gamma mixture distribution. Cell numbers were corrected for CD45$^-$CD45$^+$ FACS-sorted ratio (Supplementary Table 1) before applying negative binomial regression modelling. Modelling was achieved using the glm.nb function in the R MASS package. Modelled cell number data were studied for regression coefficient significance (variable coefficient $P \leq 0.05$) to the response variable of gestational age with the corresponding *z*-score and *P* values taken (Supplementary Table 10).

**Dimensionality reduction and trajectory analysis.** Dimensionality reduction methods included *t*-distributed stochastic neighbor embedding (*t*-SNE) (Seurat, computed from the first 20 PCs, Barnes–Hut fast computation), UMAP (Python UMAP package, five nearest neighbours, correlation metric, minimum distance 0.3, computed from the first 20 PCs), FDG (ForceAtlas2 class from fa2 Python Package, Barnes–Hut implementation for faster computation with theta 0.8, 2,000 iterations) and PAGA (paga in scanpy Python package v.1.2.2). Development trajectories were inferred by comparing FDG, PAGA and diffusion-map plots. Inferred trajectory analysis included computing diffusion map (scanpy tl.diffmap with 20 components), pseudotime (scanpy tl.dpt setting the earliest known cell type as root) and variable genes across pseudotime. The order of cells in pseudotime was statistically significant using Kruskal–Wallis test ($P < 1 \times 10^{-7}$).

Comparisons of trajectories across stage were performed by subsetting liver dataset by stage using the SubsetData function, computing dimensional reduction coordinates, batch correcting by sample using Harmony and plotting PAGA and FDG by stage. Cell-type comparisons across tissue involved subsetting for cell types of interest using the SubsetData function, merging cross-tissue datasets using the MergeSeurat function and processing data using the same approach as for the liver and NLT datasets. Harmony batch correction was then performed by tissue type, with results presented as combined UMAPs, FDGs and PAGA score heatmaps.

**Dynamically expressed genes across pseudotime.** Genes that vary across pseudotime were calculated using the DifferentialGeneTest function in Monocle in R (v.2.6.4) and a cut-off of adjusted $P$ value $<0.001$ was applied. This was applied on the entire pseudotime range and also on the pseudotime intervals specific to each cell type in order to avoid limitation to the genes characterized by monotonic changes across the inferred trajectory. Expression of pseudotime-variable genes were min − max normalized before visualization and annotated based on the involvement of each gene in relevant cell-specific functional modules or hallmark functional pathways from MSigDB v.6.2, a curated molecular signature database[60]. Peak expression for each gene over pseudotime was calculated and grouped into 'early', 'mid' or 'late' categories. For visualization purposes, the resulting gene lists were minimized by ordering them from those present in the most selected functional pathways to least, as well as ensuring coverage across pseudotime. These genes were manually compared against current literature to determine if they have known functional or cell type associations. The top 20–25 genes in each list were displayed using the ggplot2 package. Transcription factors were marked within the dataset based on AnimalTFDB transcription factor prediction database[61]. The full pseudotime gene list is available in the interactive files accompanying diffusion maps.

Mouse data generated previously[20], which were used for comparison of erythroid-associated genes expressed in mouse Kupffer cells versus EI macrophages shown in Extended Data Fig. 4g, are available at the Gene Expression Ominbus with accession code GSE127980.

**Visualization by animated FDG representation.** The FDG animation was created using an in-house modified version of the ForceAtlas2 class in fa2 Python package by saving all the intermediate states (published version only outputs the final state and discards all intermediates). The FDG coordinates at each iteration were plotted and the resulting graphs were assembled in a mp4 video format using VideoWriter in cv2 (v.3.3.1) Python package.

**Differential gene extraction and validation.** Differential gene validation was done using a random forest classifier (RandomForestClassifier class in ensemble module of sklearn Python package v.0.19.1, with 500 estimators, min_sample_split of 5, class weights set to the 'balanced' policy and all other parameters set to default). The random forest algorithm was chosen as it resembled the FACS gating hierarchy. Seventy per cent of the data was used for training and 30% for test. Parameter tuning was performed on training data using grid search. To determine whether tissue-related transcriptome variations were present in equivalent immune populations between liver, skin and kidney, each equivalent population was taken in turn and grouped according to its tissue of origin. Seurat FindMarkers function was then applied in a pairwise manner between each tissue subset to produce a cell-type-specific list of genes marking each tissue subset. These were investigated in turn for biological relevance, with representative genes displayed using the VlnPlot function of Seurat.

DEGs from B cell pseudotime were studied for significant expression change across stage and differentiation state using a one-way ANOVA with Tukey's multiple comparison test. DEGs displaying significant variance in ln-normalized expression were further studied for correlation to DEGs identified within all other cell-types across stage. B cell pseudotime DEGs with significantly correlated trends of expression to DEGs within other cell-types across stage ($P < 0.05$, two-tailed Pearson's R at 95% confidence interval) were plotted in Prism (v.8.1.2, GraphPad Software). All graphs presented in the manuscript were plotted using ggplot2 R package, Seurat implementation of ggplot2, matplotlib Python package, Prism (v.8.1.2, GraphPad Software) or FlowJo (v.10.4.1). Spot plots are shown throughout the manuscript, displaying scaled expression of ln-normalized counts.

**Primary immunodeficiency gene list curation.** Disease and genetic deficiency information was extracted from a previous study[9] and manually annotated to include HGNC symbol names for each disease-associated genetic defect for subsequent correlation with the liver dataset. Diseases implicated in primary immunodeficiency (PID) were divided according to the International Union of Immunological Societies (IUIS) major categories and screened across the liver scRNA-seq dataset. Three-hundred-and-fifteen unique genes were identified in the dataset from the 354 inborn errors of immunity highlighted in the article. For each disease category, a dot plot was generated using Seurat DotPlot function and ordered by highest expression across each gene and across each cell type, highlighting those cell types in each disease category that express the highest number of genes associated with a genetic defect.

**CellPhoneDB analysis.** CellPhoneDB v.2.0 (www.cellphonedb.org)[62] was used for the receptor–ligand analysis in Extended Data Fig. 4a. Significant ($P < 0.05$) receptor–ligand interactions between VCAM1$^{+}$ EI macrophages and the two erythroid (early and mid) populations were displayed.

**Whole-genome sequencing and fetal cell identification.** To identify maternal cells present in our data, we combined the information from fetal whole-genome DNA sequencing with the single-cell RNA-seq data. For each sample, we measured the allele frequency in the fetal DNA of SNPs from the 1,000 genomes project[63] falling within exons with a population allele frequency in excess of 1%. We then considered only those SNPs that are homozygous in the fetal DNA for follow up in the scRNA-seq data. A SNP was considered to be homozygous if its allele frequency in the fetal DNA was less than 0.2 or greater than 0.8 and had a false discovery rate-adjusted $P$ value of less than 0.01 under a binomial test for the null hypothesis that the allele frequency in the DNA was in the range [0.3,0.7].

The allele frequency of each of these SNPs with population allele frequency $>1\%$ that are known to be homozygous in the fetal DNA was then measured in each cell in the scRNA-seq data. Any deviations from homozygosity in the RNA-seq data must be a consequence of either sequencing errors, RNA editing or the genotype of the cell differing from the fetal DNA. For each cell, we calculated the total fraction of reads at the SNPs (selected as described above) that differ from the fetal genotype. We then assume that the genome-wide rate of deviations due to sequencing errors and RNA editing is less than or equal to 2%. For maternal cells, the expected genome-wide rate of deviation at these SNPs is equal to half the mean of the population allele frequency at the interrogated SNPs. Finally, for each cell we calculated the posterior probability of the cell being fetal or maternal assuming a binomial distribution with rate 2% for a fetal cell and half the mean of the population allele frequency for the maternal cell and assigned a cell as: maternal or fetal if either posterior probability exceeded 99%; ambiguous otherwise. We validated this method using samples for which both the fetal and maternal DNA were available.

**Imaging mass cytometry.** Antibodies were conjugated to metals using the Fluidigm MaxPar conjugation kits and the associated method with the following modifications: the lanthanides were used at 1.5 mM and washed for a shorter duration (4 × 5 min) in W-buffer before elution. Ultrapure MilliQ water was used throughout for any dilutions and washes. Four-micrometre-thick formalin-fixed paraffin-embedded sections obtained from 8 and 15 PCW fetal liver tissue blocks were incubated at 60 °C for 1 h then dewaxed in xylene (Fisher). After rehydration through graded alcohols (Fisher) and a 5-min wash in water, the sections were subjected to heat-induced epitope retrieval with citrate buffer (pH 6.0). Sections were then washed in water and PBS (Gibco) and blocked with 3% BSA (Sigma-Aldrich) for 45 min. A mixture of eight metal-conjugated antibodies diluted in 0.5% BSA (see Supplementary Table 21 for antibody details), was added to the sections for overnight incubation at 4 °C in a humidified chamber. Slides were washed twice in 0.2% Triton X-100 diluted in PBS for 8 min and then twice in PBS for 8 min.

To counterstain nucleated cells, sections were incubated with 312.5 nm (193 Ir) Intercalator-Ir (Fluidigm) for 30 min at room temperature. Slides were then washed in water for 5 min, and allowed to air dry at room temperature before imaging on the Hyperion imaging mass cytometer. Using expected target cell frequencies from previous fluorescence flow cytometry data, region-of-interest (ROI) size was set to 2.8 mm × 3.8 mm. The ablation energy was set at 2 dB with a laser frequency of 200 Hz. Each session of ablation generated a .mcd image file containing information for every panorama and ROI measured whereby each 1-μm piece of tissue liberated by the laser was analysed for ionic content on a per channel basis by time of flight. Single-cell segmentation and feature extraction was performed using CellProfiler (v.3.1.5). Nuclei were identified using the IdentifyPrimaryObjects module, in which the input images were the sum of the DNA-stained iridium channels (191 and 193) constructed by the ImageAfterMath module. The diameter range set for nuclei identification was 4–15 pixel units. The ExpandOrShrink module was used to grow the nuclear segmentation area by three pixels to define the cellular area and the MeasureObjectIntensity module was used to determine the mean intensity for each cell object identified.

**Light-sheet and confocal fluorescence microscopy.** Male embryos at 5, 7 and 11 PCW that were deemed to be devoid of morphological anomalies were dissected after overnight fixation in 4% PFA. Whole-mount and cryosection immunostaining were performed as described previously[64], with the following conditions: tissue was incubated with primary antibodies (see Supplementary Table 22 for antibody details) for 9 days at 37 °C, with secondary antibody for 16 h at 37 °C using dedicated host species antibodies and reagent combination. TO-PRO-3 647 was used at 1:100 in whole embryos and 1:5,000 on cryosections. Whole-mount specimens were solvent-cleared as described[64], and imaged in dibenzylether with a Miltenyi Lavision Biotech ultramicroscope (Olympus MXV10 stereomicroscope and PCO Edge SCMOS CCD camera using the dedicated Inspector pro acquisition software. Four lasers (at 488, 561, 647 and 790 nm wavelengths) were used to generate light sheets. IMaris (v.9.2, BitPlane) was used for image conversion and processing. To assess co-labelling at a cellular level the DBE-cleared specimens were further imaged using an upright confocal microscope Olympus FV-1000 with the following wavelengths: 405 (diode), 440, 488 (Argon), 515 (Argon), 559 (diode) and 635 nm (diode), acquired using Olympus Fluoview 4.2.b and displayed in Imaris 9.2.4. Deconvolution was performed using NIH-FIJI. Photoshop (Adobe) was used to create panels. All raw files are available on demand through our dedicated online platform (www.transparent-human-embryo.com).

**Statistics and reproducibility.** For all analyses of fetal liver 3′ 10x data, $n = 14$ biologically independent samples were included. This includes Figs. 1b, 2a, 3a, b, 4b, 5a, b, 6a, b, g, Extended Data Figs. 1c, 2a, 3a–e, h, 4a, 5e, g–i, 6a–e, 7b, h, 8.

For all analyses of fetal liver 3′ 10x data by developmental stage, $n = 4$ 7–8 PCW, $n = 4$ 9–11 PCW, $n = 3$ 12–14 PCW and $n = 3$ 15–17 PCW biologically independent samples were used. This includes Figs. 1c, 4a, Extended Data Figs. 1d, e, 5a–c, f.

For analyses including 10x sequencing data of skin, kidney, yolk sac, decidua and placenta, $n = 7$, 3, 3, 11 and 5 biologically independent samples were used, respectively. These analyses are shown in Figs. 3b, 4b, 5a, b, 6g, Extended Data Figs. 1a, b, c–e, h, 5g–i, 6c–e.

For analysis including sequencing data of cord blood and adult bone marrow, $n = 8$ biologically independent samples of each were used. These analyses are shown in Fig. 6g and Extended Data Fig. 7h.

For all scRNA-seq data shown, all cells of a given label from indicated tissues are shown, and no downsampling or subsetting was performed. The following cell numbers generated using scRNA-seq are displayed in each of the listed figures: Fig. 1b, c, 113,063 fetal liver cells; Fig. 2a, 113,063 fetal liver cells; Fig. 3a, 104,515 fetal liver cells; Fig. 3b, 43,507 fetal liver cells, 2,455 yolk sac cells and 243 skin cells; Fig. 4a, 16,919 fetal liver cells; Fig. 4b, 16,919 fetal liver cells, 2,757 skin cells, 213 kidney cells and 259 yolk sac cells; Fig. 5a, b, 38,464 fetal liver cells, 6,887 yolk sac cells and 10,008 cells from decidua and placenta; Fig. 6a, b, 5,673 fetal liver cells; Fig. 6g, 3,439 fetal liver HSC/MPP, 205 yolk sac progenitors, 1,082 cord blood HSCs and 3,668 adult bone marrow HSCs; Extended Data Fig. 1a, 10,258 skin cells and 1,795 kidney cells; Extended Data Fig. 1b, 10,071 yolk sac cells; Extended Data Fig. 1c–e, 113,063 fetal liver cells; Extended Data Fig. 1f, 1,206 fetal liver cells from $n = 2$ biologically independent samples profiled using Smart-seq2 (also displayed in Extended Data Fig. 2b); Extended Data Fig. 2a, 113,063 fetal liver cells; Extended Data Fig. 2b, 1,206 fetal liver cells from $n = 2$ biologically independent samples profiled using Smart-seq2 and 381 fetal liver erythroid liver EI populations (early, mid and late erythroids, VCAM1$^+$ EI macrophages), Kupffer cells and endothelium validated by Smart-seq2 from $n = 2$ biologically independent fetal liver samples; Extended Data Fig. 2d, 381 fetal liver erythroid liver EI populations (early, mid and late erythroids, VCAM1$^+$ EI macrophages), Kupffer cells and endothelium validated by Smart-seq2 from $n = 2$ biologically independent fetal liver samples; Extended Data Fig. 3a, b, 52,237 fetal liver cells; Extended Data Fig. 3c, 52,237 fetal liver cells, 362 skin cells and 28 kidney cells; Extended Data Fig. 3d, 52,327 fetal liver cells, 362 skin cells, 28 kidney cells and 2,793 yolk sac cells were analysed, and a maximum of 20 cells displayed of each cell type per tissue; Extended Data Fig. 3e, 52,237 fetal liver cells, 362 skin cells, 28 kidney cells and 2,588 yolk sac cells; Extended Data Fig. 3h, 3,439 HSC/MPP, 1,342 MEMP, 11,985 early erythroid, 27,000 mid erythroid, 3,180 late erythroid, 3,983 megakaryocytes and 1,308 mast cells from fetal liver, 55 MEMP, 51 mid erythroid, 137 late erythroid, 11 megakaryocytes and 108 mast cells from skin, and 2 MEMP and 26 megakaryocytes from kidney; Extended Data Fig. 5a, 16,919 fetal liver cells; Extended Data Fig. 5b, c, 767 fetal liver early lymphoid/T lymphocytes; Extended Data Fig. 5d, 16,666 fetal liver cells from $n = 7$ biologically independent samples; Extended Data Fig. 5e, 7,467 fetal liver cells; Extended Data Fig. 5f, 32,308 fetal liver cells; Extended Data Fig. 5g, 16,919 fetal liver cells, 2,775 skin cells, 213 kidney cells and 464 yolk sac cells; Extended Data Fig. 5h, 6,706 NK cells and 1,726 ILC precursors from fetal liver, 1,479 NK cells and 1,142 ILC precursors from skin, and 155 NK cells and 36 ILC precursors from kidney; Extended Data Fig. 5i, 1,726 fetal liver ILC precursors, 1,142 skin ILC precursors and 36 kidney ILC precursors; Extended Data Fig. 6a, 6,606 fetal liver cells; Extended Data Fig. 6b, 11,653 fetal liver cells; Extended Data Fig. 6c, 38,646 fetal liver cells, 6,887 yolk sac cells and 10,008 cells from decidua and placenta were analysed, and a maximum of 50 cells displayed of each cell type per tissue; Extended Data Fig. 6d, 43,498 fetal liver cells, 8,350 skin cells, and 1,514 kidney cells; Extended Data Fig. 6e, 24,841 Kupffer cells, 2,586 monocytes, 253 pDC precursors, 336 DC1 and 3,954 DC2 from fetal liver; 5,474 macrophages, 704 monocytes, 36 pDCs, 99 DC1 and 527 DC2 from skin; and 1,075 macrophages, 82 monocytes, 1 pDC, 1 DC2 and 271 DC2 from kidney; Extended Data Fig. 7b, 6,606 fetal liver cells and 161 of 349 scRNA-seq profiled cells (Supplementary Table 12) from $n = 3$ biologically independent samples, sorted as per FACS gates in Extended Data Fig. 7a, h, 3,439 HSC/MPPs, 1,341 MEMPs, 234 pre pro-B cells, 658 neutrophil-myeloid progenitors, 350 monocyte precursors, 253 pDC precursors and 330 DC precursors from fetal liver, 205 yolk sac progenitors; 1,082 cord blood HSCs; and 3,668 adult bone marrow HSCs; Extended Data Fig. 8, 113,063 fetal liver cells.

Representative mini bulk RNA-seq data of 100 cells per cell state from 1 of $n = 3$ biologically independent fetal liver samples are shown in Extended Data Fig. 2b, c. Cytospin images shown in Fig. 2b and Extended Data Fig. 5e are representative from 1 of $n = 3$ biologically independent samples. Hyperion images shown in Extended Data Fig. 2e are representative from 1 of $n = 4$ biologically independent 8 PCW fetal livers and 1 of $n = 4$ biologically independent 15 PCW fetal livers. Immunohistochemical (IHC) staining of 8 PCW fetal skin in Extended Data Fig. 3f is representative from 1 of $n = 3$ biologically independent samples. IHC staining of fetal liver in Extended Data Fig. 4b are from independent samples and both are representative from 1 of $n = 3$ biologically independent samples. Light-sheet fluorescence microscopy of embryo (5 PCW) hand skin shown in Extended Data Fig. 3g is a representative image from 1 of $n = 3$ biologically independent samples.

Statistical analysis of differential gene expression was done using one-tailed Wilcoxon rank sum test with Bonferroni correction, including those shown in heatmaps and violin plots. In Fig. 6b, expression of genes in HSC/MPP 1 compared to each other cell cluster, and MEMP/neutrophil-myeloid progenitor/pre pro-B cell compared to each other cell cluster was statistically significant unless specified as not significant (NS). All $P < 0.001$, except expression of *KLF1* between HSC/MPP1 vs HSC/MPP2 ($P = 0.00934$), *IL1RL1* between HSC/MPP 1 vs HSC/MPP 2 ($P = 0.00148$), *GATA2* between MEMP vs HSC/MPP2 ($P = 0.00802$), *BCL11A* between HSC/MPP 1 vs HSC/MPP 7 ($P = 0.00114$), *LYZ* between HSC/MPP 1 vs HSC/MPP 5 ($P = 0.00675$), *AZU1* between HSC/MPP 1 vs HSC/MPP 8 ($P = 0.00494$) and *CSF1R* between neutrophil-myeloid progenitors vs HSC/MPP 8 ($P = 0.213$). In Fig. 6g, ****$P < 0.0001$ when comparing expression of *ANXA1*, *DUSP1*, *HLA-B* and *HSPAIA* between samples. Expression of genes in fetal liver early lymphoid/T lymphocytes was compared across developmental stages and displayed in Extended Data Fig. 5b, c. No significant difference across stage was observed in the genes displayed in Extended Data Fig. 5b. The following comparisons shown in Extended Data Fig. 5c were significant: *CD2* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$) and vs 12–17 PCW ($P < 0.0001$); *TRDC* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$) and 9–11 PCW vs 12–17 PCW ($P < 0.0001$); *CD8A* expression at 7–8 PCW vs 9–17 PCW ($P = 0.00714$); *CD27* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$) and 9–11 PCW vs 12–17 PCW ($P < 0.0001$); *IL7R* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$), 9–11 PCW vs 12–17 PCW ($P = 0.00168$), and 12–14 PCW vs 15–17 PCW ($P = 0.00216$); *JCHAIN* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$); *CD3D* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$); *KLRB1* expression at 9–11 PCW vs 12–17 PCW ($P < 0.0001$); *TRAC* expression at 7–8 PCW vs 9–17 PCW ($P < 0.0001$) and 9–11 PCW vs 12–17 PCW ($P < 0.0001$); and *PRF1* expression at 9–11 PCW vs 12–17 PCW ($P < 0.0001$). Statistically significant gene expression shown in Extended Data Fig. 5h comparing expression in NK cells and ILC precursors across tissues was $P < 0.001$ where **** was shown, and $P = 0.00236$ for *TXNIP* expression in ILC precursors in liver compared to those in kidney. Expression of genes in ILC precursors in fetal liver, skin and kidney, as shown in Extended Data Fig. 5i, was compared between tissues and no significant difference was observed. Extended Data Fig. 6e shows comparisons between fetal liver, skin and kidney for macrophages, monocytes, pDCs, DC1 and DC2. The following comparisons were statistically significant ($P < 0.0001$) between macrophages from different tissues: *CD14* expression in kidney vs liver and vs skin; *CD68* expression in liver vs skin and vs kidney; *CD163*, *RNASE1* and *F13A1* expression between all tissues; and *VCAM1* expression in liver vs skin and vs kidney. The following comparisons were statistically significant between monocytes from different tissues: *CD14* expression in liver vs skin ($P < 0.0001$); *CD68* expression in liver vs skin ($P < 0.0001$); *S100A9* expression in liver vs kidney ($P < 0.0001$), and skin vs kidney ($P = 0.0245$); *FCGR3A*. expression in liver vs skin ($P = 0.0004$); *POSTN* expression in liver vs skin ($P < 0.0001$) and vs kidney ($P < 0.0001$), and skin vs kidney ($P = 0.0411$). No significant difference was observed when comparing genes in pDCs across tissues. *CLEC9A* expression was statistically significantly different in liver vs skin DC1 ($P < 0.0001$). The following comparisons were statistically significant between DC2 from different tissues: *CD1C* expression in skin vs liver ($P < 0.0001$) and vs kidney ($P < 0.0001$); *CLEC10A* expression in kidney vs liver ($P < 0.0001$) and vs skin ($P < 0.0001$); *S100B* expression in skin vs liver ($P < 0.0001$) and vs kidney ($P = 0.0162$); and *FCER1A* and *CD83* expression between each tissue ($P < 0.0001$).

Statistical analysis of HSC colony assays shown in Fig. 6c–e and Extended Data Fig. 7 was done by applying two-tailed Fisher's exact tests to colony counts. Figure 6c, d and Extended Data Fig. 7f show 125, 217 and 124 colonies from 7–8 PCW, 12–14 PCW and 15–17 PCW fetal liver samples, respectively, from $n = 2$ biologically independent samples per development stage. The number of colonies per sample is 93, 32, 93, 124, 84 and 40, which are shown in Extended Data Fig. 7c. The number of colonies between the following stages in Fig. 6c were statistically significant: erythroid colonies in 7–8 PCW vs 15–17 PCW ($P = 0.0238$), erythroid/megakaryocyte/myeloid colonies in 7–8 PCW vs 15–17 PCW ($P = 0.0294$), NK colonies in 7–8 PCW vs. 15–17 PCW ($P = 0.0357$), and erythroid/myeloid colonies in 7–8 PCW vs 12–14 PCW ($P = 0.0188$) and 15–17 PCW ($P < 0.001$), and 12–14 PCW vs 15–17 PCW ($P = 0.0232$). The number of erythroid-containing colonies was significant between each stage shown in Fig. 6d ($P < 0.001$). The number of colonies that differentiated along three lineages was significant between 7–8 PCW vs 12–14 PCW ($P = 0.0041$), and 7–8 PCW vs 15–17 PCW ($P = 0.0027$). Figure 6e shows 141, 74 and 124 colonies from 7–8 PCW, 12–14 PCW and 15–17 PCW fetal liver samples, respectively, from $n = 2$ biologically independent samples per development stage. The number of B cell-forming colonies in Fig. 6e was significant between 7–8 PCW vs 12–14 PCW ($P = 0.0014$) and 15–17 PCW ($P = 0.0044$). Extended Data Fig. 7d, e shows 163, 196 and 182 colonies from $n = 3$ 7–8 PCW, $n = 2$ 12–14 PCW and $n = 2$ 15–17 PCW biologically independent fetal liver samples, respectively on the left, and 42, 74, 47, 97, 99, 59 and 123 colonies by individual sample on the right. The number of erythroid colonies compared to all other types shown in Extended Data Fig. 7d was statistically significant between 12–14 PCW and 15–17 PCW ($P = 0.0307$). The number of erythroid-containing colonies was significant between 7–8 PCW vs 15–17 PCW ($P = 0.0013$), and 12–14 PCW vs 15–17 PCW ($P = 0.0497$), as shown in Extended Data Fig. 7e. Extended Data Fig. 7f shows 125, 217 and 124 colonies from 7–8 PCW, 12–14 PCW and 15–17 PCW fetal liver samples, respectively, from $n = 2$ biologically independent samples per development stage. Extended Data Fig. 7g shows 141, 74 and 124 colonies from 7–8 PCW, 12–14 PCW and 15–17 PCW fetal liver samples, respectively, from $n = 2$ biologically independent samples per development stage. The percentage of NK-containing colonies was statistically significant in 7–8 PCW vs 15–17PCW ($P = 0.0032$), and 12–14 PCW vs 15–17 PCW ($P = 0.0115$).

Flow cytometric analysis of cell cycle phases, as shown in Fig. 6g and Extended Data Fig. 7i was performed on cells from $n = 3$ 7–8 PCW and $n = 3$ 12–16 PCW biologically independent fetal liver samples, and $n = 2$ biologically independent cord blood samples. One-way ANOVA with Tukey's multiple comparison test was used to determine statistical significance between stages (7–8 PCW and 12–14 PCW) and samples (fetal liver and cord blood). The percentage of CD34$^+$CD38$^-$ cells in G0 was significantly higher in 12–14 PCW livers compared with 7–8 PCW livers ($P = 0.0136$).

Cell cycle phases determined by transcriptome analysis of fetal liver cells, fetal skin cells, and fetal kidney cells are shown in Extended Data Fig. 3h. Statistical significance of the proportion of MEMPs and megakaryocytes in each cell cycle phase was compared between fetal liver, skin and kidney using a Kruskal–Wallis test with Dunn's post hoc test. Statistical significance of the proportion of mid and late erythroids, and mast cells in each cell cycle phase was compared between fetal liver and skin using two-tailed Mann–Whitney tests. The following comparisons were statistically significant: megakaryocytes in fetal liver vs fetal kidney in G1 ($P = 0.0317$), G2M ($P = 0.0317$) and S ($P = 0.0139$); megakaryocytes in fetal liver vs fetal skin in S ($P = 0.0039$); mid erythroids in fetal liver vs skin in G1 ($P = 0.0031$) and G2M ($P < 0.0001$); late erythroids in fetal liver vs fetal skin in G1 ($P = 0.0021$) and S ($P < 0.0001$); and mast cells in fetal liver vs fetal skin in G1 ($P = 0.0248$) and S ($P = 0.0337$).

Statistical comparison of the percentage of MEMPs, mid and late erythroids, megakaryocytes and mast cells expressing *MKI67* in fetal liver vs NLT (skin and kidney) was performed using two-tailed Mann–Whitney tests. This is displayed in Extended Data Fig. 3h, where the following comparisons were statistically significant: megakaryocytes in fetal liver vs fetal NLT ($P = 0.007$), mid erythroid in fetal liver vs fetal skin ($P = 0.0305$), and late erythroid in fetal liver vs fetal skin ($P = 0.0368$).

Statistical comparison of the percentage of fetal liver HSC/MPP expressing *MKI67* compared to the percent of fetal liver MEMP, pre pro-B cells, neutrophil-myeloid progenitors, monocyte precursors, pDC precursors and DC precursors, yolk sac progenitors, and HSC from cord blood and adult bone marrow expressing *MKI67* was performed using a Kruskal–Wallis test with Dunn's post hoc test. This is displayed in Extended Data Fig. 7h, in which the percentage of *MKI67*-expressing fetal liver HSC/MPPs was statistically significant when compared with the following populations: fetal liver MEMPs ($P = 0.0180$), monocyte precursors ($P < 0.0001$), DC precursors ($P < 0.0001$), cord blood HSC ($P < 0.0001$) and adult bone marrow HSC ($P = 0.0076$).

Extended Data Fig. 1g displays a flow cytometric analysis of the frequency of B cells in the CD34$^-$ cells from $n = 13$ 6–9 PCW, $n = 13$ 9–12 PCW, $n = 14$ 12–15 PCW and $n = 9$ 15–19 PCW biologically independent fetal liver samples. Statistical significance across the developmental stages was compared using a Kruskal–Wallis test with Dunn's multiple comparison post hoc test, for which the following comparisons were statistically significant: 6–9 PCW vs 12–15 PCW ($P < 0.0001$), 6–9 PCW vs 15–19 PCW ($P = 0.0003$), 9–12 PCW vs 12–15 PCW ($P = 0.0157$) and 9–12 PCW vs 15–19 PCW ($P = 0.0287$).

Statistically significant, dynamically variable genes from pseudotime were determined by likelihood of ratio test applied in monocle (see Methods). Select genes were displayed for erythroid, mast cell and megakaryocyte lineages shown in Extended Data Fig. 3b, B cell lineage shown in Extended Data Fig. 5e, and DC1, DC2 and monocyte lineages shown in Extended Data Fig. 6b.

Investigation of potential receptor–ligand interactions between all fetal liver VCAM1$^+$ EI macrophages (161 cells) and early and mid erythroids (11,985 and 27,000 cells, respectively), as shown in Extended Data Fig. 4a, was performed using CellPhoneDB. A permutation test was applied to determine statistical significance, which is indicated by the colour of the dots. ln-normalized median expression of *ITGA4* and *VCAM1* in the same cells is also displayed in a violin plot.

The ImageStream analysis of fetal liver cells shown in Extended Data Fig. 4c–f was performed on $n = 3$ biologically independent samples. Extended Data Fig. 4c, d display representative data from one sample. Extended Data Fig. 4e shows representative images of cells from 38,576 single cells and 1,945 doublets from one sample. The percentage of each cell type that was observed within the doublet of singlet gate, as per Extended Data Fig. 4c, was compared using a Kruskal–Wallis test with Dunn's post hoc test. The percentage of doublets in VCAM1$^+$ EI macrophages was significantly different from the percentage of doublets in erythroids ($P = 0.0194$).

Statistical significance of *TNFSF13B* expression in Kupffer cells over time compared to *NFKBIA* in HSC/MPPs ($P = 0.0245$), pre pro-B cells ($P = 0.0008$), pro-B cells ($P = 0.0004$), pre-B cells ($P = 0.0197$) and B cells ($P = 0.0343$) across four developmental stages spanning 6–17 PCW was determined using two-tailed Pearson's R test.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The raw sequencing data, expression count data with cell classifications are deposited at ArrayExpress with accession code E-MTAB-7407.

## Code availability

All scripts are available at https://github.com/haniffalab/FCA_liver.

53. Gerrelli, D., Lisgo, S., Copp, A. J. & Lindsay, S. Enabling research with human embryonic and fetal tissue resources. *Development* **142**, 3073–3076 (2015).
54. Bullen, P. & Wilson, D. in *Molecular Genetics of Early Human Development* (eds Strachan, T. et al.) 27–35 (Bios Scientific, 1997).
55. Hern, W. M. Correlation of fetal age and measurements between 10 and 26 weeks of gestation. *Obstet. Gynecol.* **63**, 26–32 (1984).
56. Roy, A. et al. High resolution IgH repertoire analysis reveals fetal liver as the likely origin of life-long, innate B lymphopoiesis in humans. *Clin. Immunol.* **183**, 8–16 (2017).
57. Villani, A.-C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017).
58. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
59. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
60. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
61. Hu, H. et al. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38 (2019).
62. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB v2.0: Inferring cell-cell communication from combined expression of multi-subunit receptor-ligand complexes. Preprint at *bioRxiv* https://doi.org/10.1101/680926 (2019).
63. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
64. Belle, M. et al. Tridimensional visualization and analysis of early human development. *Cell* **169**, 161–173 (2017).

**Acknowledgements** This publication is part of the Human Cell Atlas (www.humancellatlas.org/publications). We thank the Newcastle University Flow Cytometry Core Facility, Bioimaging Core Facility, Genomics Facility, NUIT for technical assistance, School of Computing for access to the High-Performance Computing Cluster and Newcastle Molecular Pathology Node Proximity Laboratory; A. Farnworth for clinical liaison, S. Hambleton for primary immunodeficiency expertise, H. Chen for immunohistochemistry assistance and M. Belle and S. Fouquet for light-sheet fluorescence microscopy assistance. The human embryonic and fetal material was provided by the Joint MRC–Wellcome (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org). We acknowledge funding from the Wellcome Human Cell Atlas Strategic Science Support (WT211276/Z/18/Z); M.H. is funded by Wellcome (WT107931/Z/15/Z), The Lister Institute for Preventive Medicine and NIHR and Newcastle Biomedical Research Centre; S.A.T. is funded by Wellcome (WT206194), ERC Consolidator and EU MRG-Grammar awards; S.B. is funded by Wellcome (WT110104/Z/15/Z) and St. Baldrick's Foundation; E.L. is funded by a Wellcome Sir Henry Dale and Royal Society Fellowship, European Haematology Association, Wellcome and MRC to the Wellcome–MRC Cambridge Stem Cell Institute and BBSRC; A. Regev is funded by the Manton Foundation and Klarman Cell Observatory.

**Author contributions** M.H., S.A.T. and E.L. conceived and directed the study. M.H., S.A.T., E.L. and R.A.B. designed the experiments. Samples were collected by S. Lisgo and S. Lindsay, isolated by R.A.B. and libraries prepared by E.S., L.M., D.-M.P., R.V.-T., J.-E.P. and J.F. Flow cytometry and FACS experiments were performed by R.A.B., E.F.C., L.J., D. Maunder. and A. Filby. Imaging mass cytometry experiments were performed by M.A., B.M., B.I., D. McDonald and A. Fuller. Cytospins were performed by D.D. and J.F. and in vitro culture differentiation experiments were performed by L.J., D. Maunder. and E.F.C. Immunohistochemistry was performed by B.I., M.A., F.G. and C.M.; P.C. and M.A. interpreted immunohistochemistry and developmental pathology sections. Y.G. and A.C. performed and interpreted light-sheet fluorescence microscopy experiments. M.S.K., B.L., O.A., M.T., D.D., T.L.T., M.S., O.R.-R. and A. Regev. generated adult and cord blood scRNA-seq datasets. D.-M.P., K.G., K.P., S.W., I.G., M.E., P.V., M.D.Y., Z.M. and J.B performed the computational analysis. M.H., D.-M.P., R.A.B., B.G., E.L, I.R., A. Roy, E.F.C., L.J., A.-C.V., R.R., E.P., M.M., J.-E.P., G.R., K.B.M., M.J.T.S., A. Filby., K.G., S.W., I.G., S.B. and J.B. interpreted the data. M.H., L.J., R.A.B., E.S., D.-M.P., B.G., E.L., I.R., K.G., S.W., I.G., A.-C.V. and A. Roy. wrote the manuscript. All authors read and accepted the manuscript.

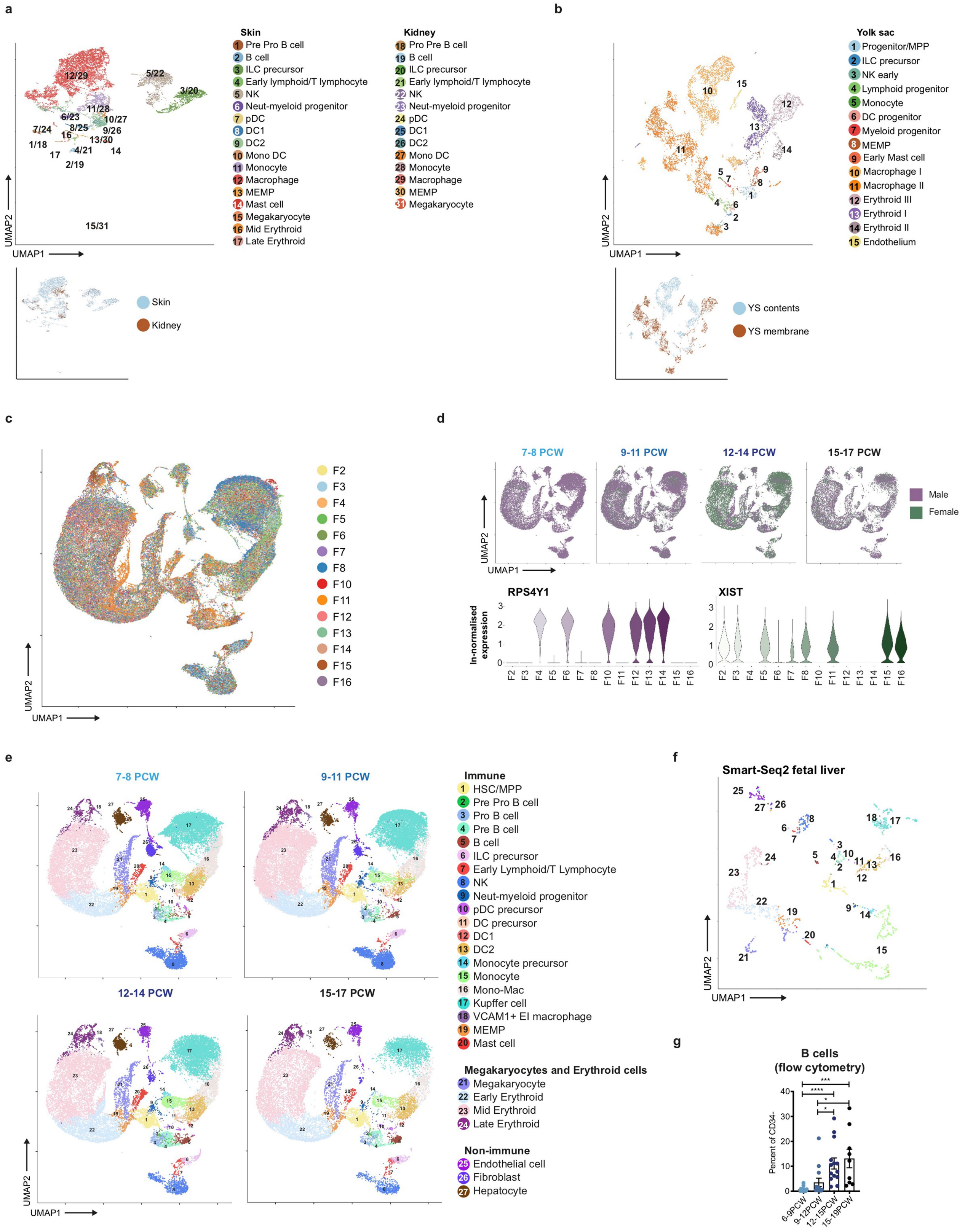
**Competing interests** A. Regev is a founder and equity holder of Celsius Therapeutics and an SAB member of Neogene Therapeutics, ThermoFisher Scientific and Syros Pharmaceuticals.

**Additional information**

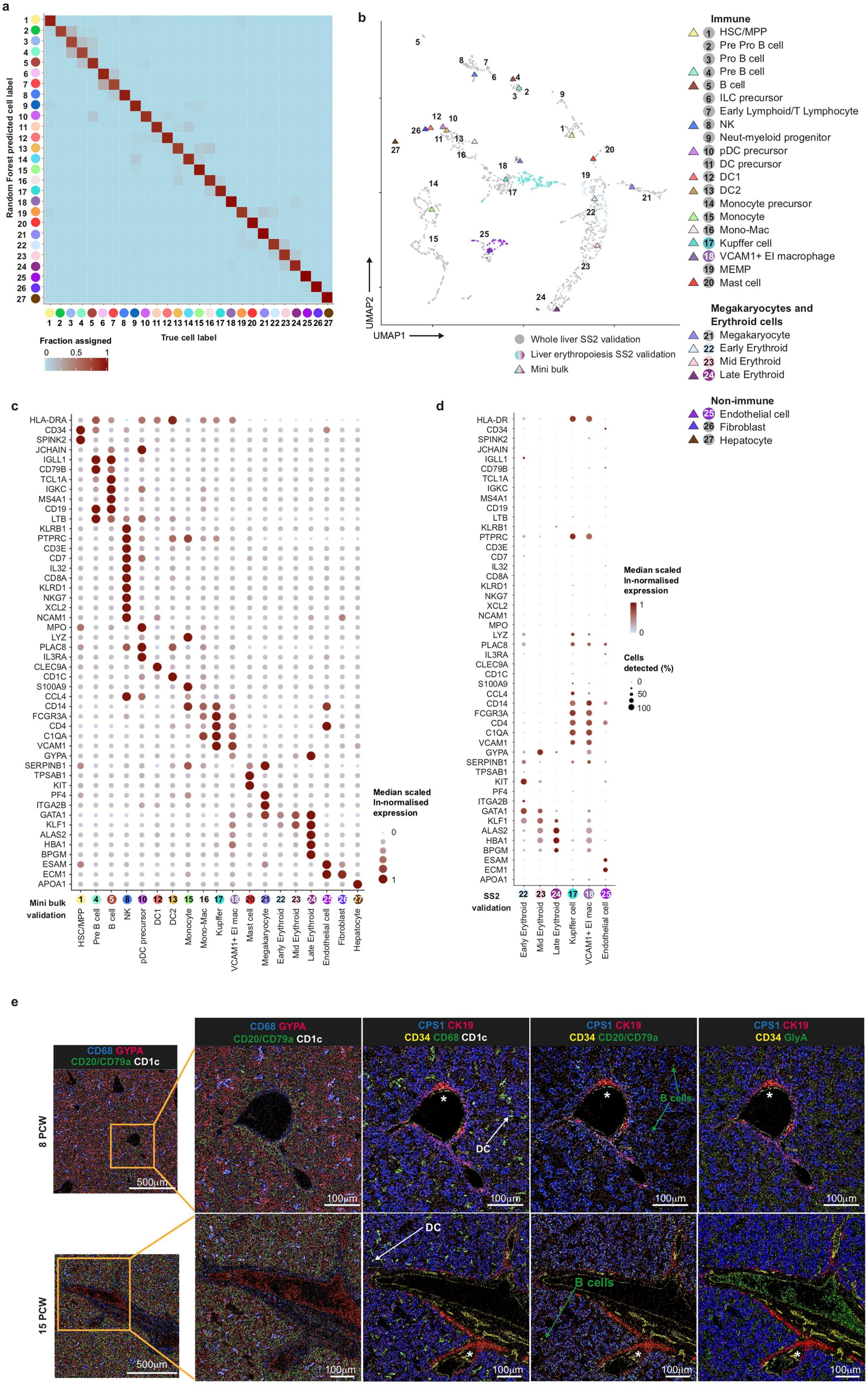
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1652-y.

**Correspondence and requests for materials** should be addressed to S.B. or E.L. or S.A.T. or M.H.

**Reprints and permissions information** is available at http://www.nature.com/reprints.

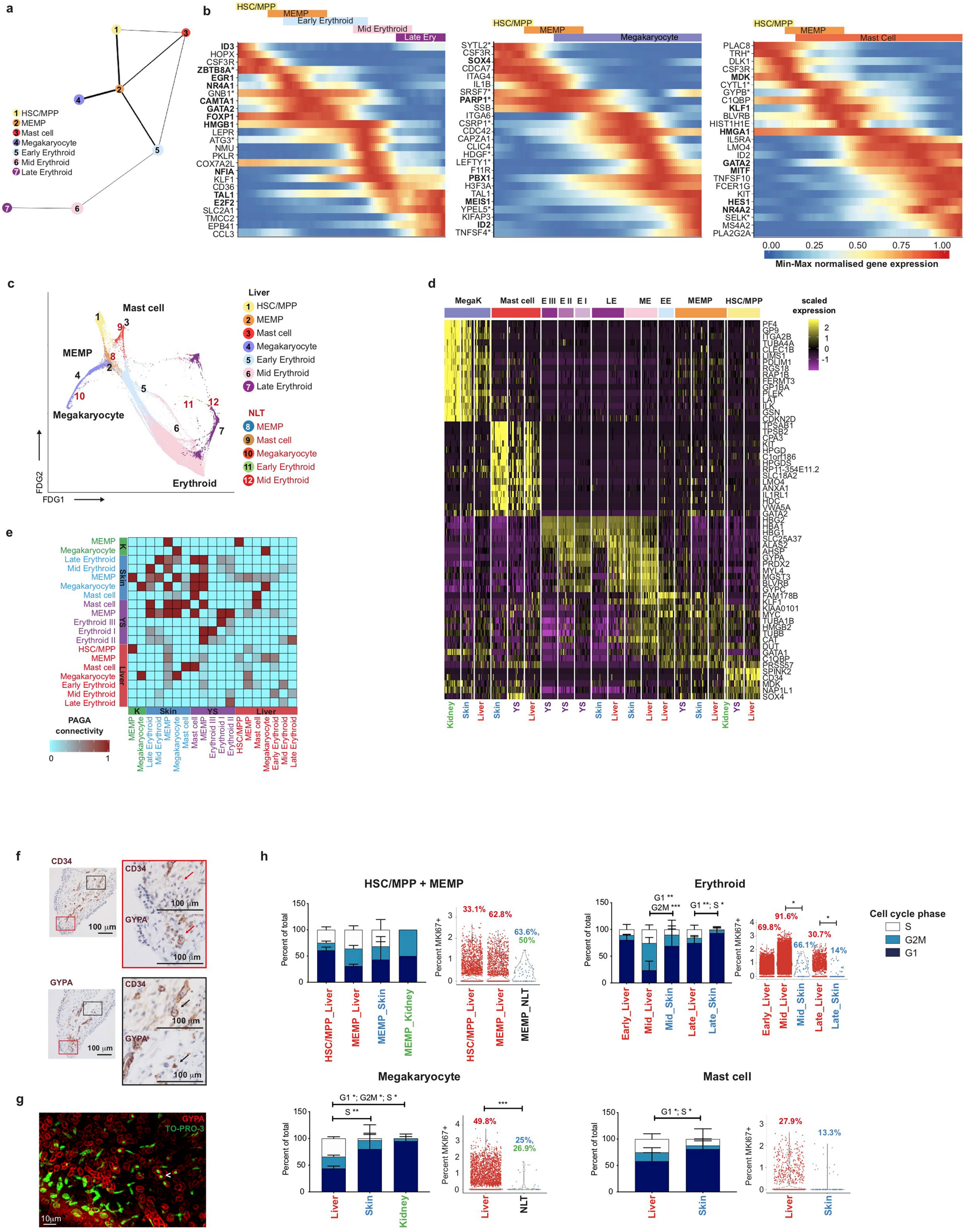


**Extended Data Fig. 1 | Single-cell transcriptome map of fetal liver.** **a**, Fetal skin and kidney haematopoietic cells visualized by UMAP. Colours indicate cell state. Inset, colours indicate tissue type. **b**, UMAP visualization of yolk sac haematopoietic cells. Colours indicate cell state. Inset, colours indicate location within yolk sac. **c**, UMAP visualization of 3′ liver 10x cells after batch correction, coloured by sample. **d**, UMAP visualization (top) of 3′ 10x liver sample sex mixing grouped by developmental stage, and violin plots (bottom) showing ln-normalized median expression of *XIST* (green) and *RSP4Y1* (purple), which mark female and male samples, respectively. **e**, UMAP visualization of fetal liver composition by developmental stage. Colours indicate cell state. **f**, UMAP visualization of fetal liver cells profiled using Smart-seq2. Colours indicate cell states as shown in **e**. **g**, Frequency (mean ± s.e.m.) of B cells in the CD34$^{-}$ cells detected in 6–19 PCW fetal livers by flow cytometry ($^{*}P < 0.05$, $^{***}P = 0.003$ and $^{****}P < 0.001$).



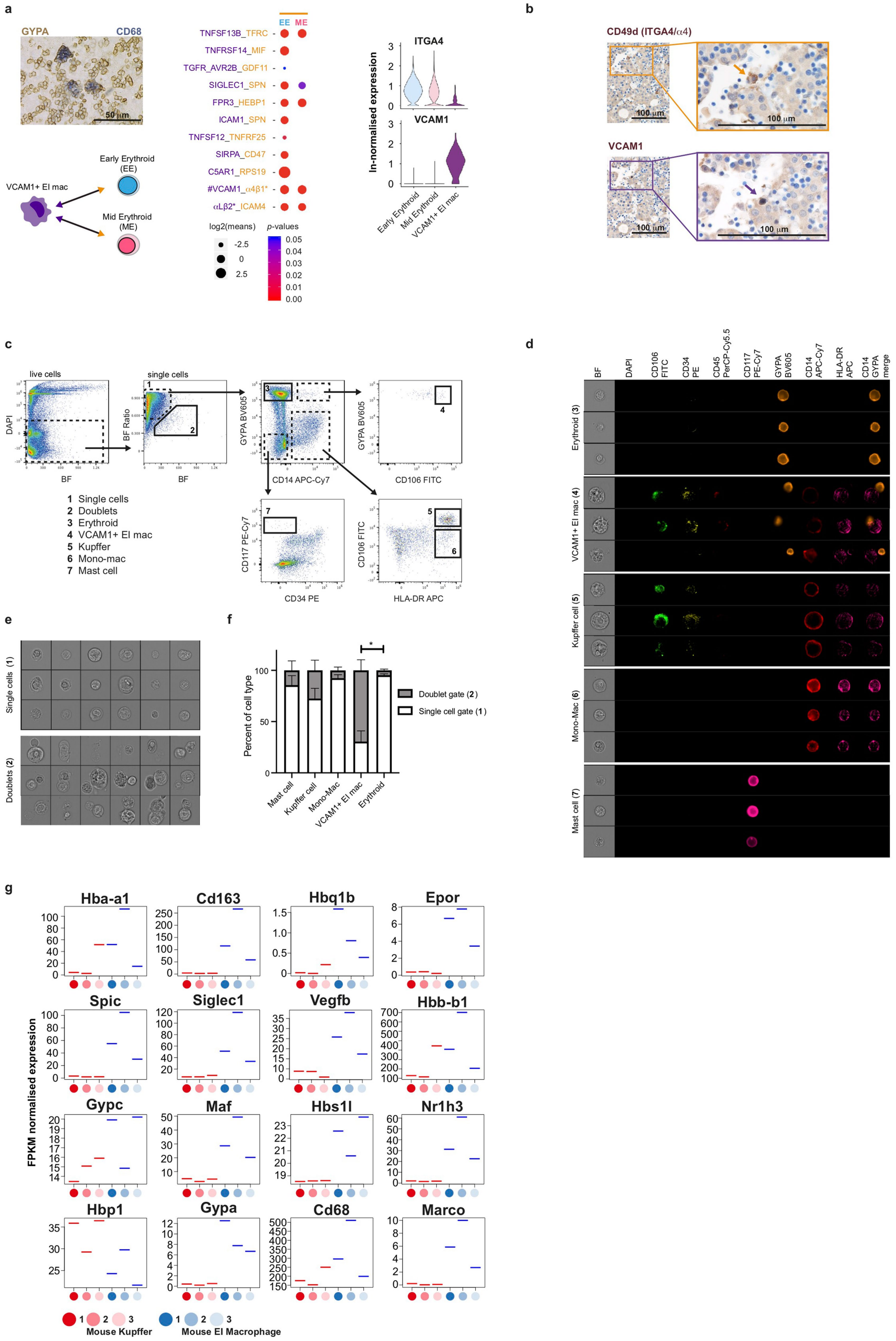
**Extended Data Fig. 2** | See next page for caption.

**Extended Data Fig. 2 | Transcriptome validation of fetal liver cells.** **a**, Assessment of 48 genes from the 4,471 highly variable genes by using a random forest classifier to assign cell labels, where 'true cell label' indicates the manual annotation based on the full list of variable genes. **b**, Comparison of representative mini bulk RNA-seq data (in coloured triangles) and liver EI populations (early, mid and late erythroids, VCAM1$^+$ EI macrophages), Kupffer cells and endothelium validated by Smart-seq2 (SS2) (in colour) overlaid on whole liver SS2 populations (grey). **c**, Dot plot showing representative median-scaled ln-normalized gene expression of 100 FACS-isolated liver cells based on marker gene expression in Fig. 2a. Gene expression indicated by spot size and colour intensity. **d**, Dot plot showing median-scaled ln-normalized gene expression of FACS-sorted single cells from liver EI populations (early, mid and late erythroids, VCAM1$^+$ EI macrophages), Kupffer cells and endothelium shown as coloured dots in **b** based on marker gene expression in Fig. 2a. Gene-expression frequency (per cent of cells within cell type expressing the gene) indicated by spot size and expression level by colour intensity. **e**, Overlay pseudo-colour Hyperion representative images for 8 PCW and 15 PCW fetal liver. Far left images are shown at 5× magnification with zoom of insets on right at 20× magnification (1 μm per pixel). Asterisks indicate bile ducts.



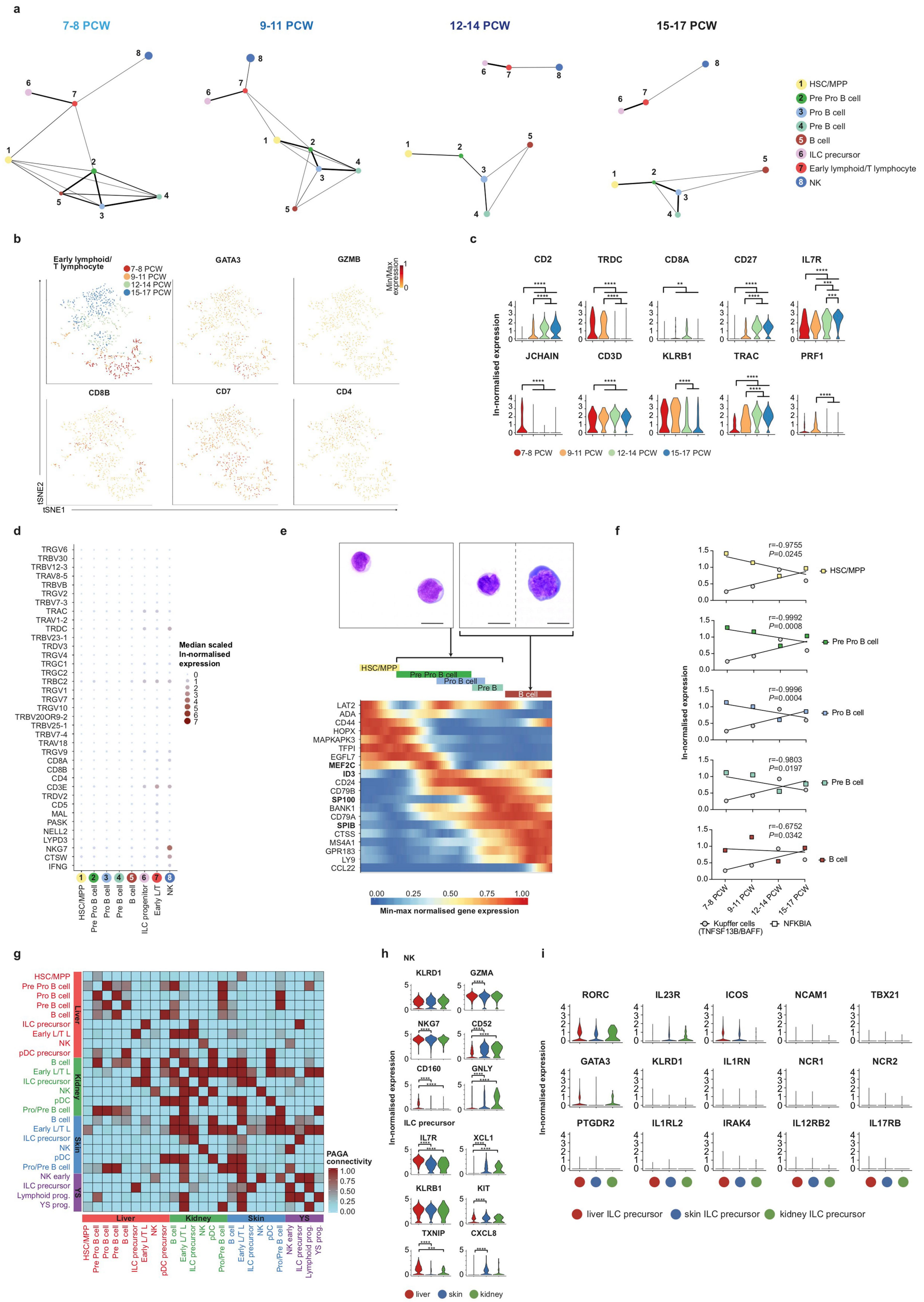
**Extended Data Fig. 3 |** See next page for caption.

**Extended Data Fig. 3 | Fetal liver and NLT haematopoiesis. a**, PAGA analysis of fetal liver HSC/MPP, erythroid, megakaryocyte and mast cell lineages from Fig. 3a. Lines show connections; line thickness corresponds to the level of connectivity (low (thin) to high (thick) PAGA connectivity). **b**, Heat map showing min − max normalized expression of statistically significant ($P < 0.001$), dynamically variable genes from pseudotime analysis for erythroid, megakaryocyte and mast cell inferred trajectories. Transcription factors in bold, asterisks mark genes not previously implicated for the respective lineages. **c**, FDG visualization of fetal liver, skin and kidney HSC/MPP, MEMP, erythroid, megakaryocyte and mast cell lineages. **d**, Heat map showing the scaled ln-expression of selected marker genes in fetal liver, NLT and yolk sac subsets. **e**, PAGA connectivity scores of HSC/MPP, erythroid, megakaryocyte and mast cell lineages between fetal liver, skin, kidney (K) and yolk sac. **f**, Representative immunohistochemical staining of sequential sections of 8 PCW fetal skin for endothelium (CD34$^+$) and erythroblasts (nucleated and GYPA$^+$), nuclei stained with blue alkaline phosphatase. Zoom in of insets (right) bordered with black (top) indicate nucleated cells stained positive for GYPA within CD34$^+$ blood vessels, and those bordered with red (bottom) indicate nucleated GYPA$^+$ cells outside CD34$^+$ blood vessels. Scale bars, 100 μm. **g**, Representative confocal fluorescence microscopy of embryo (5 PCW) hand skin. Scale bar, 10 μm; red, TO-PRO-3 nuclei; green, GYPA (see also Supplementary Video 2 showing light-sheet fluorescence microscopy). The arrowhead indicates extravascular nucleated erythroid cells. **h**, Stacked bar plots (left) showing percentage (mean ± s.d.) of fetal liver (red), skin (blue) and kidney (green) HSC/MPP, MEMP, erythroid, megakaryocyte and mast cells in each stage of the cell cycle (G1 (navy), G2M (blue) and S (white) phase), and ln-normalized median expression of *MKI67* transcript (right) in corresponding liver vs NLT cell types (total percent of *MKI67*-expressing cells shown above plots; each dot represents a single cell). $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.005$.



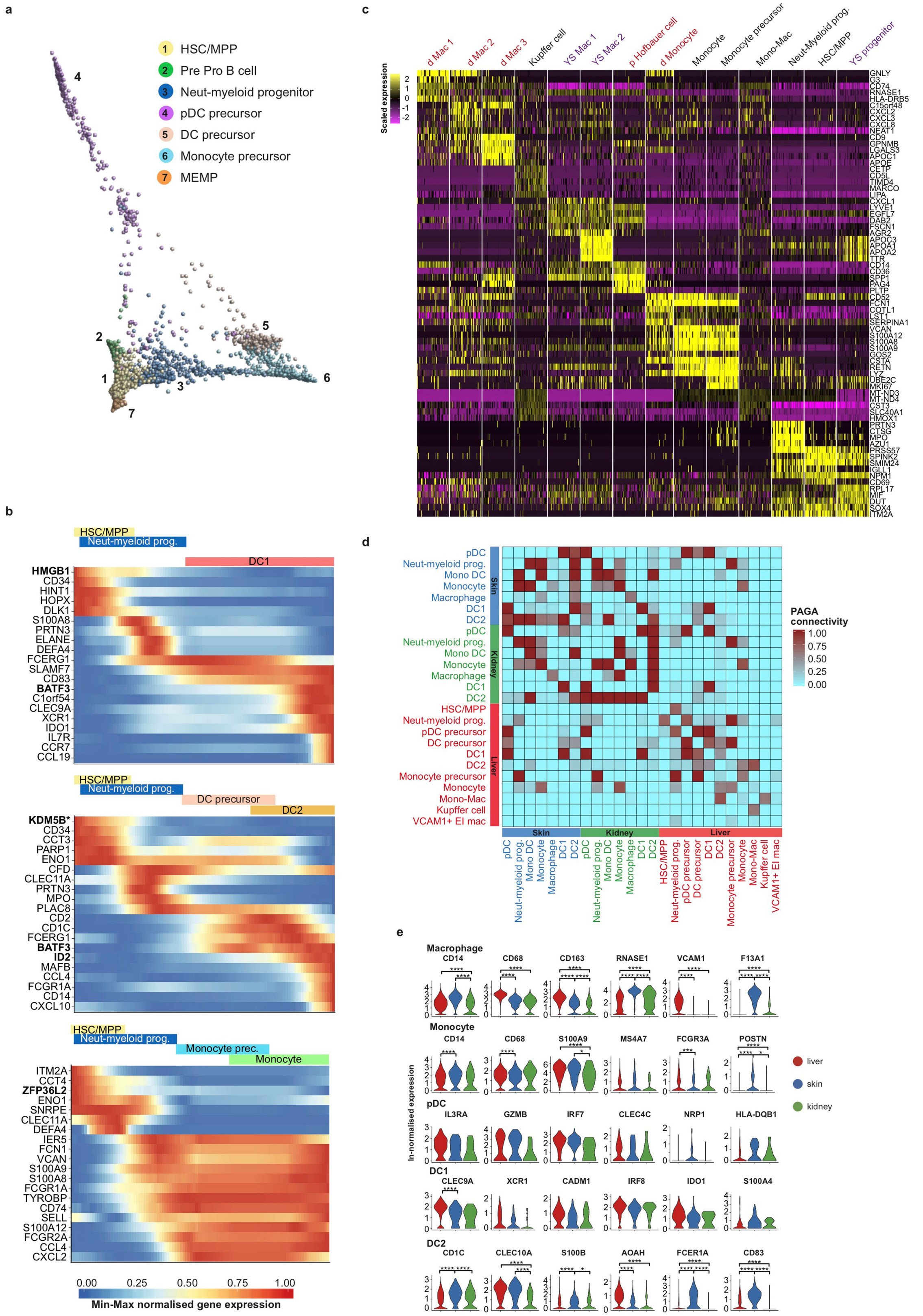
**Extended Data Fig. 4 |** See next page for caption.

**Extended Data Fig. 4 | Investigation of interactions between fetal liver macrophages and erythroid cells. a**, Representative immunohistochemical staining of fetal liver for erythroblasts and macrophages with GYPA and CD68, respectively. Scale bar, 50 μm. Statistically significantly ($P < 0.05$) enriched receptor–ligand interactions from CellPhoneDB between VCAM1$^{+}$ EI macrophages (purple) and two erythroid populations (early and mid; red) ($n = 14$ biologically independent samples). Asterisks indicate protein complexes. Violin plots show ln-normalized median gene-expression value of *VCAM1* and *ITGA4* in cells analysed by CellPhoneDB (indicated by # in dot plot). **b**, Representative immunohistochemical staining of sequential sections of 8 PCW fetal liver for VCAM1$^{+}$ EI macrophages (VCAM1$^{+}$) and CD49d$^{+}$GYPA$^{+}$ cells with nuclei stained using blue alkaline phosphatase. Right, zoom in of insets, with coloured arrows indicating erythroblast (yellow) and VCAM1$^{+}$ EI macrophage (purple). Scale bars, 100 μm. **c**, Representative gating strategy used to visualize fetal liver erythroid cells, VCAM1$^{+}$ EI macrophages, Kupffer cells, monocyte-macrophages and mast cells. **d**, Bright field (BF), VCAM1 (CD106), CD34, CD45, KIT (CD117), GYPA, CD14 and HLA-DR images for each cell type within gates shown in **c**. **e**, Representative bright-field images of cells found within the single-cell and doublet gates. **f**, Bar plots showing the proportion of each cell type within the single-cell gate (white) or doublet gate (grey) (mean ± s.d.); $*P = 0.0194$. **g**, Comparison of macrophage and erythroid gene expression in mouse macrophages (red) and EI macrophages (blue), $n = 3$ from ref. [20].

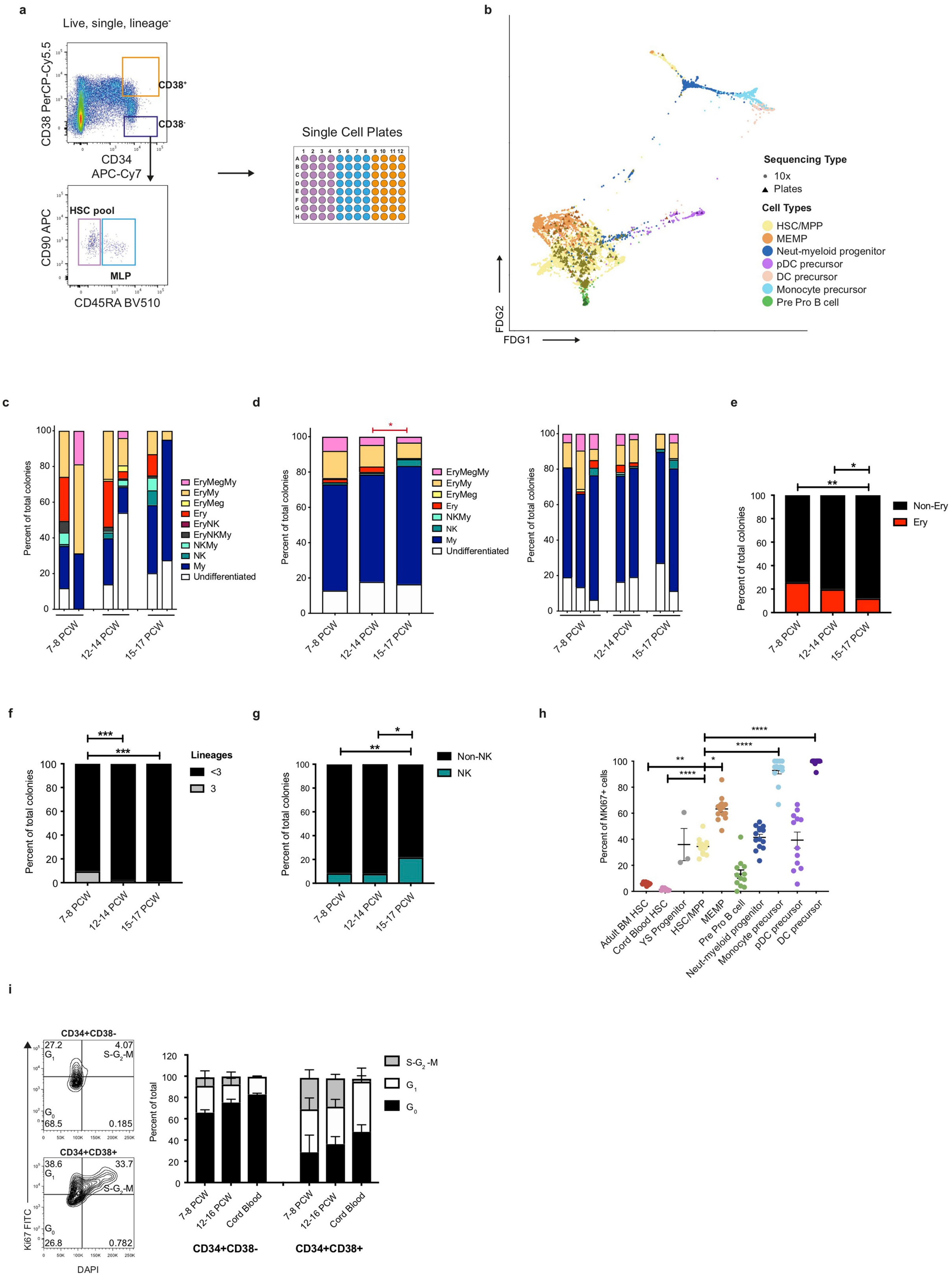


**Extended Data Fig 5 |** See next page for caption.

**Extended Data Fig 5 | Lymphoid lineages in fetal liver and NLT.** **a**, PAGA analysis of fetal liver HSC/MPP and lymphoid cell types from Fig. 1b showing changes over four developmental stages. Lines symbolize connection; line thickness corresponds to the level of connectivity (low (thin) to high (thick) PAGA connectivity). **b**, **c**, Feature plots (**b**) and violin plots (**c**) showing ln-normalized median expression of selected known NK, ILC and T cell genes over gestation for early lymphoid/T lymphocyte cluster; $^{**}P < 0.001$, $^{***}P < 0.005$ and $^{****}P < 0.0001$. **d**, Dot plot showing median-scaled ln-normalized median expression of V(D)J transcripts in fetal liver lymphoid cell types. Gene expression indicated by spot size and colour intensity. **e**, Heat map showing normalized expression of statistically significant, dynamically variable genes from pseudotime analysis for B cell lineage inferred trajectory (likelihood ratio test). Transcription factors are in bold. Morphology of liver pro-B and pre-B cells and B cells by Giemsa stain after cytospin. **f**, ln-normalized expression (mean ± s.e.m.) of *TNFSF13B* in Kupffer cells and *NFKBIA* in HSC/MPPs and cells in the B cell lineage across four developmental stages spanning 6–17 PCW; trend lines showing linear regression. **g**, PAGA connectivity scores of HSC/MPP and lymphoid cells from fetal liver, skin, kidney and yolk sac. **h**, ln-normalized median expression of selected known NK (top) and ILC precursor (bottom) marker genes and selected DEGs between liver (red), skin (blue) and kidney (green) visualized by violin plots ($^{***}P < 0.005$ and $^{****}P < 0.001$). **i**, Violin plots showing ln-normalized median expression of selected known ILC and NK cell genes expressed in ILC precursors from fetal liver, skin and kidney.



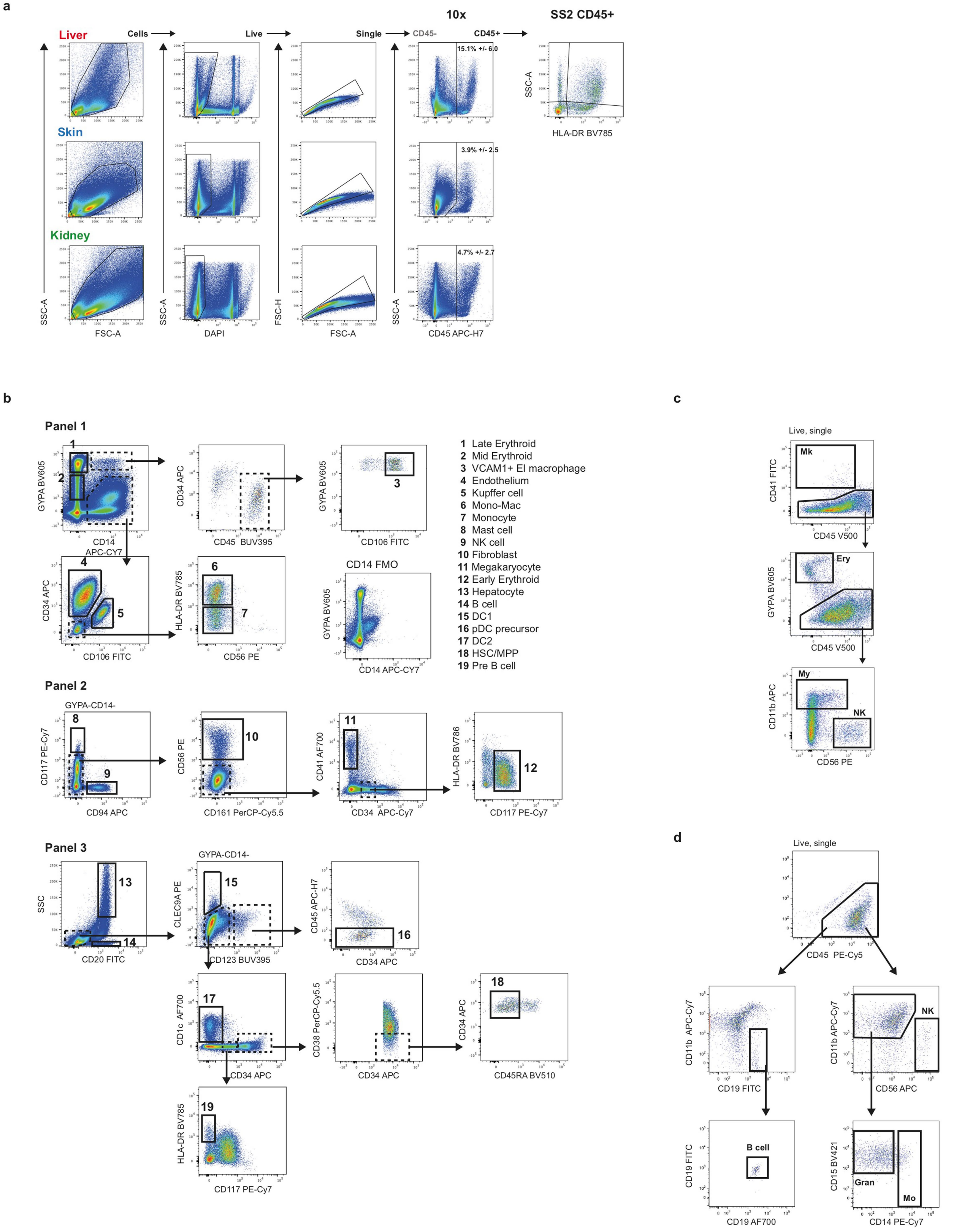
**Extended Data Fig. 6 | Tissue signatures in developing myeloid cells.** **a**, Diffusion map of fetal liver HSC/MPP, progenitors and precursors from Fig. 1b. **b**, Heat map showing min − max normalized expression ($P < 0.001$) of dynamically variable genes from pseudotime analysis for monocyte, DC1 and DC2 inferred trajectories. Transcription factors in bold, asterisks mark genes not previously implicated for the respective lineages. **c**, Heat map visualization comparing scaled expression of the top marker genes of decidua–placenta (red), fetal liver (black) and yolk sac (purple) progenitor and myeloid populations. **d**, PAGA connectivity scores of HSC/MPP and myeloid cells from fetal liver, skin and kidney. **e**, ln-normalized median expression of three known marker genes and three DEGs in corresponding myeloid populations across fetal liver, skin and kidney visualized by violin plots (*$P < 0.05$, ***$P < 0.005$ and ****$P < 0.0001$).



**Extended Data Fig. 7** | See next page for caption.

**Extended Data Fig. 7 | HSC/MPP differentiation potential by gestation.** **a**, Experimental design for single-cell transcriptome and culture of fetal liver cells from the representative FACS gates illustrated. **b**, Alignment of 349 scRNA-seq-profiled cells from FACS gates in **a** with 10x-profiled HSC/MPPs and early progenitors visualized using FDG; point shape corresponds to sequencing type (triangle, SS2 plate data; circle, 10x data). **c**, Stacked bar plot of all different types of colonies generated by single HSC pool gate cells (gate defined in **a**). **d**, Stacked bar plot of all different types of colonies generated by single HSC pool gate cells without MS5 stroma layer (gate defined in **a**) by stage (left) and in individual samples (right), $*P < 0.05$. **e**, Percentage of colonies generated by single HSC pool cells without MS5 stroma layer containing erythroid cells (sum of ery, ery–meg, ery–meg–my and ery–my colonies shown in **c**); $**P < 0.01$. **f**, Percentage of colonies from single-cell culture (shown in Fig. 6c) that differentiated along three lineage (defined as sum of ery, NK and my, and ery, meg and my colonies) branches ($***P < 0.005$). **g**, Percentage of colonies containing NK cells following B/NK optimized culture of ten cells from the HSC pool gate ($*P < 0.05$ and $**P < 0.01$). **h**, Percentage (mean ± s.e.m.) of HSC/MPPs and early progenitors in fetal liver, yolk sac, cord blood and adult bone marrow expressing *MKI67* ($*P < 0.05$, $**P < 0.01$ and $****P < 0.001$). **i**, Percentage (mean ± s.e.m.) of $CD34^+CD38^-$ and $CD34^+CD38^+$cells in the indicated cell cycle phases (right) as determined by flow cytometry analysis (left; representative plot of $n = 8$ biologically independent samples) (G0, $Ki67^-DAPI^-$; G1, $Ki67^+DAPI^-$; S–G2–M, $Ki67^+DAPI^+$ (left)).

**Extended Data Fig. 8 | Expression of known PID-linked genes in fetal liver.** Dot plots showing relative expression of genes known to be associated with major PID disease categories in fetal liver cell types from Fig. 1b. Gene-expression frequency (per cent of cells within the cell type expressing the gene) is indicated by spot size and expression level is indicated by colour intensity.



**Extended Data Fig. 9 | FACS gating strategy for scRNA-seq analysis.** **a**, Gating strategy used to isolate cells for droplet (10x) and plate-based scRNA-seq (Smart-seq2) for samples F2–F17. **b**, Gating strategy used to isolate cells for cytospins, scRNA-seq (Smart-seq2) and 100-cell RNA-seq. **c**, Flow cytometry gating strategy used to identify the colonies cultured in vitro from single cells as shown in Fig. 6c. **d**, Flow cytometry gating strategy used to identify B and NK colonies cultured in vitro from 10 cells as shown in Fig. 6e.

nature research

Corresponding author(s): Muzlifah Haniffa, Sarah Teichmann, Sam Behjati & Elisa Laurenti

Last updated by author(s): Aug 8, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

Data collection

No software was used for data collection.

Data analysis

Alignment, quantification and quality control of scRNA-seq data
Droplet-based (10x) sequencing data was aligned and quantified using the Cell Ranger Single-Cell Software Suite (version 2.0.2, 10x Genomics Inc) using the GRCh38 human reference genome (official Cell Ranger reference, version 1.2.0). Smart-seq2 sequencing data was aligned with STAR (version 2.5.1b), using the STAR index and annotation from the same reference as the 10x data. Gene-specific read counts were calculated using htseq-count (version 0.10.0). Cells with fewer than 200 detected genes and for which the total mitochondrial gene expression exceeded 20% were removed. Genes that were expressed in fewer than 3 cells were also removed. We detected on average ~3,000 genes per cell with the 10x Genomics platform and ~6,000 genes with the Smart-seq2 protocol.

Doublet detection
Doublets were detected with an approach adapted from Pijuan-Sala et al.. In the first step of the process, each 10x lane was processed independently. Scrublet was run, obtaining per-cell doublet scores. The standard Seurat-inspired Scanpy processing pipeline was performed up to the clustering stage, using default parameters. Each cluster was subsequently separately clustered again, yielding an over clustered manifold, and each of the resulting clusters had its Scrublet scores replaced by the median of the observed values. The resulting scores were assessed for statistical significance, with p-values computed using a right-tailed test from a normal distribution centred on the score median and a MAD-derived standard deviation estimate. The MAD was computed from above-median values to circumvent zero-truncation. The p-values were FDR-corrected with the Benjamini-Hochberg procedure, and a significance threshold of 0.1 was imposed. In the second step of the process, all 10x lanes for a single tissue were pooled together and the Seurat-inspired Scanpy processing was repeated, with the addition of Harmony7 with a theta of 3 for batch correction between the lanes before the neighbour graph identification step. The joint manifold was clustered, and the frequency of identified doublets was computed. The same statistical framework as in the first step was used to identify clusters significantly enriched in doublets, which were subsequently flagged as doublets in their entirety and removed.

Clustering and annotation

Downstream analysis included data normalisation (NormalizeData, LogNormalize method, scaling factor 10000), data feature scaling (ScaleData), variable gene detection (FindVariableGenes), PCA (RunPCA, from variable genes) and Louvain graph-based clustering (FindClusters, data dimensionality reduction using PCA, clustering resolution (res.30)) performed using the R package Seurat (version 2.3.4). Cluster cell identity was assigned by manual annotation using known marker genes and computed differentially expressed genes (DEGs) using FindAllMarkers function in Seurat package (Wilcoxon rank sum test, p-values adjusted for multiple testing using the Bonferroni correction; Supplementary Table 3). For computing DEGs all genes were probed provided they were expressed in at least 25% of cells in either of the two populations compared and the expression difference on a natural log scale was at least 0.25. Manual annotation was performed iteratively, which included validating proposed cell labels with known markers and further investigating clusters whose gene signatures indicated additional diversity.

Clustering and cell type assignment for fetal liver data was assessed using two additional clustering methods (not shown): Agglomerative clustering (with Ward linkage and Euclidean affinity) and Gaussian mixture (AgglomerativeClustering class from cluster module and GaussianMixture from mixture module in sklearn version 0.19.1 Python 3.6.3). Consensus agreement between the 3 clustering methods was measured by Rand index and adjusted mutual information implemented in the metrics module in sklearn package. The Rand Index scores were 0.89 and 0.85 for Agglomerative and Gaussian Mixture clustering methods respectively.

After annotation was completed, a cell type classifier was built by training an SVM on labelled fetal liver scRNA-seq data with grid search for parameter optimization based on training data. 70% of the data was used for training and the other 30% for test. The SVM was previously compared in terms of accuracy and recall with a random forest and logistic regression classifiers trained on the same data. Out of the 3 classifiers the SVM was chosen due to showing a mean accuracy and weighted mean recall of 95%. Random forest showed 89% for both precision and recall (Supplementary Table 7). The SVM classifier was used for automatic annotation of the Smart-seq2 and mini bulk RNA sequencing data sets to allow identification of biologically meaningful clusters and DEG computation.

Data generated from fetal skin, kidney and yolk sac was pre-processed, normalised, clustered and manually annotated, in parallel with, and using the same pipeline as, the liver data. Skin and kidney data were combined using the MergeSeurat function. Clusters characterised by differentially expressed immune gene markers were extracted from the NLT dataset for subsequent comparative analysis with liver-derived immune populations. Human cord blood and adult bone marrow datasets were downloaded from Human Cell Atlas data portal (https://preview.data.humancellatlas.org/). These were processed using the same approach as described above, followed by manual annotation. Decidua and placental data from Vento-Tormo et al., were downloaded from ArrayExpress record E-MTAB-6701 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6701/).

Data integration
We used Harmony data integration to correct for batch effect between sample identities. The average kBET rejection rate statistically significantly improved from 0.735 to 0.471 (Supplementary Table 8) following Harmony data integration (p-value 3.83e-3 in Kolmogorov–Smirnov test and p-value 8.8e-6 in Wilcoxon signed-rank test). The manifold was subjected to re-clustering using Harmony adjusted PCs with parameters as mentioned above in “Clustering and annotation”. Cell type classifications were then ascertained through re-annotation of the clusters derived from Harmony adjusted PCs to produce the final annotation.

Changes in cell proportions over development
Comparison of cell proportions across gestational stages was assessed using negative binomial regression on cell number data using the glm.nb function in the R MASS package. Cell numbers were corrected for CD45-/CD45+ FACS sorted ratio (Supplementary Table 1) prior to applying negative binomial regression.

Dimensionality reduction and trajectory analysis
Dimensionality reduction methods included tSNE (Seurat, computed from the first 20 PCs, Barnes-Hut fast computation), UMAP (Python UMAP package, 5 nearest neighbours, correlation metric, minimum distance 0.3, computed from the first 20 PCs), FDG (ForceAtlas2 class from fa2 Python Package, Barnes-Hut implementation for faster computation with theta 0.8, 2000 iterations) and partition-based approximate graph abstraction (PAGA) (paga in scanpy Python package version 1.2.2). Development trajectories were inferred by comparing FDG, PAGA and diffusion map plots. Inferred trajectory analysis included computing diffusion map (scanpy tl.diffmap with 20 components), pseudotime (scanpy tl.dpt setting the earliest known cell type as root) and variable genes across pseudotime. Order of cells in pseudotime was statistically significant using Kruskal-Wallis test (p < 1x10-7).

Comparisons of trajectories across stage were performed by subsetting liver dataset by stage using SubsetData function, computing dimensional reduction coordinates, batch correcting by sample using Harmony, and plotting PAGA and FDG by stage. Cell type comparisons across tissue involved subsetting for cell types of interest using SubsetData function, merging cross-tissue datasets using MergeSeurat function, and processing data using the same approach as for the liver and NLT datasets. Harmony batch correction was then performed by tissue type, with results presented as combined UMAPs, FDGs and PAGA score heatmaps.

Dynamically expressed genes across pseudotime
Genes that vary across pseudotime were calculated using DifferentialGeneTest function in Monocle in R (version 2.6.4) and a cut-off of adjusted p-value < 0.001 applied. This was applied on the entire pseudotime range and also on the pseudotime intervals specific to each cell type in order to avoid limitation to the genes characterised by monotonic changes across the inferred trajectory. Expression of pseudotime variable genes were min-max normalized prior to visualization and annotated based on each gene’s involvement in relevant cell-specific functional modules or hallmark functional pathways from MSigDB v6.2, a curated molecular signature database66. Peak expression for each gene over pseudotime was calculated and grouped into ‘Early’, ‘Mid’ or ‘Late’ categories. For visualization purposes, the resulting gene lists were minimized by ordering them from those present in the most selected functional pathways to least, as well as ensuring coverage across pseudotime. These genes were manually compared against current literature to determine if they have known functional or cell type associations. The top 20-25 genes in each list were displayed using the ggplot2 package. Transcription factors were marked within the dataset based on AnimalTFDB transcription factor prediction database. The full pseudotime gene list is available in the interactive files accompanying diffusion maps.

Visualization by animated force-directed graph representation
The FDG animation was created using an in-house modified version of the ForceAtlas2 class in fa2 Python package by saving all the intermediate states (published version only outputs the final state and discards all intermediates). The FDG coordinates at each iteration were plotted and the resulting graphs were assembled in a mp4 video format using VideoWriter in cv2 (version 3.3.1) Python package.

Differential gene extraction and validation
Differential gene validation was done using a random forest classifier (RandomForestClassifier class in ensemble module of sklearn Python package v0.19.1, with 500 estimators, min_sample_split of 5, class weights set to the "balanced" policy and all other parameters set to default). The Random Forest algorithm was chosen as it resembled the FACS gating hierarchy. 70% of the data was used for training and 30% for test. Parameter tuning was performed on training data using grid search. To determine whether tissue-related transcriptome variations were present in equivalent immune populations between liver, skin and kidney, each equivalent population was taken in turn and grouped according to its tissue of origin. Seurat FindMarkers function was then applied in a pair-wise manner between each tissue subset to produce a cell type-specific list of genes marking each tissue subset. These were investigated in turn for biological relevance, with representative genes displayed using VlnPlot function of Seurat.

DEGs from B-cell pseudotime were studied for significant expression change across stage and differentiation state using a one-way ANOVA with Tukey's multiple comparison test. DEGs displaying significant variance in ln-normalised expression were further studied for correlation to DEGs identified within all other cell-types across stage. B-cell pseudotime DEGs with significantly correlated trends of expression to DEGs within other cell-types across stage (p-value <0.05, Two-tailed Pearson's R at 95% CI) were plotted in Prism (v8.1.2, GraphPad Software). All graphs presented in the manuscript were plotted using ggplot2 R package, Seurat implementation of ggplot2, matplotlib Python package, Prism (v.8.1.2, GraphPad Software) or FlowJo (v10.4.1). Spot plots are shown throughout the manuscript, displaying scaled expression of ln-normalized counts.

Primary immunodeficiency (PID) gene list curation
Disease and genetic deficiency information was extracted from Picard et al. and manually annotated to include HGNC symbol names for each disease-associated genetic defect for subsequent correlation with the liver dataset. Diseases implicated in PID were divided according to the International Union of Immunological Societies (IUIS) major categories and screened across the liver scRNA-seq dataset. 315 unique genes were identified in the dataset from the 354 inborn errors of immunity highlighted in the article. For each disease category a dot plot was generated using Seurat DotPlot function and ordered by highest expression across each gene and across each cell type, highlighting those cell types in each disease category which express the highest number of genes associated with a genetic defect.

CellPhoneDB analysis
CellPhoneDB v2.0 (www.cellphonedb.org) was used for the receptor-ligand analysis in Figure 3d. Significant ($p < 0.05$) receptor-ligand interactions between VCAM1+ Erythroblastic Island macrophages and the two erythroid (early and mid) populations were displayed.

Whole genome sequencing and fetal cell identification
To identify maternal cells present in our data we combined the information from fetal whole genome DNA sequencing with the single cell RNA-seq data. For each sample we measured the allele frequency in the fetal DNA of SNPs from the 1000 genomes project falling within exons with a population allele frequency in excess of 1%. We then consider only those SNPs which are homozygous in the fetal DNA for follow up in the scRNA-seq data. A SNP was considered to be homozygous if its allele frequency in the fetal DNA was less than 0.2 or greater than 0.8 and had an FDR adjusted p-value of less than 0.01 under a binomial test for the null hypothesis that the allele frequency in the DNA was in the range [0.3,0.7].

The allele frequency of each of these SNPs with population allele frequency > 1% that are known to be homozygous in the fetal DNA was then measured in each cell in the scRNA-seq data. Any deviations from homozygosity in the RNA-seq data must be a consequence of either sequencing errors, RNA editing, or the genotype of the cell differing from the fetal DNA. For each cell, we calculated the total fraction of reads at the SNPs (selected as described above) that differ from the fetal genotype. We then assume that the genome-wide rate of deviations due to sequencing errors and RNA editing is less than or equal to 2%. For maternal cells, the expected genome wide rate of deviation at these SNPs is equal to half the mean of the population allele frequency at the interrogated SNPs. Finally, for each cell we calculated the posterior probability of the cell being fetal or maternal assuming a binomial distribution with rate 2% for a fetal cell and half the mean of the population allele frequency for the maternal cell and assign a cell as: maternal/fetal if either posterior probability exceeds 99%, ambiguous otherwise. We validated this method using samples for which both the fetal and maternal DNA were available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data, expression count data with cell classifications are deposited at ArrayExpress: experiments: E-MTAB-7407.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Determined based on prior literature and Human Cell Atlas recommendations as per Human Cell Atlas white paper (www.humancellatlas.org/news/13) |
| Data exclusions | Any exclusions are stated in the quality control checks for single cell RNA-seq data in Methods. |
| Replication | Similar results were obtained using orthogonal methodological approaches including experiments performed across two laboratories. |
| Randomization | Samples were randomly allocated to this study through HDBR |
| Blinding | N/A |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | Antibodies |
| ☐ | ☒ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology |
| ☒ | ☐ | Animals and other organisms |
| ☐ | ☒ | Human research participants |
| ☒ | ☐ | Clinical data |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | ChIP-seq |
| ☐ | ☒ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

## Antibodies

Antibodies used

All antibody details (clone and manufacturer) are included in Supplementary methods (Tables 8-14) and also listed below:
Antibody Clone Manufacturer
CD45 PerCP-Cy5.5 2D1 BD Bioscience
HLA-DR APC L243 BD Bioscience
CD34 PE 8G12 BD Bioscience
CD45 APC-H7 2D1 BD Bioscience
HLA-DR BV785 L243 Biolegend
CD14 APC-Cy7 MφP9 BD Bioscience
CD34 APC 581 BD Bioscience
CD45 BUV395 HI30 BD Bioscience
CD56 PE NCAM16.2 BD Bioscience
CD106 FITC 51-10C9 BD Bioscience
GYPA BV605 GA-R2 BD Bioscience
CD14 PE CF594 MφP9 BD Bioscience
CD34 APC-Cy7 581 Biolegend
CD41 AF700 HIP8 Biolegend
CD56 PE NCAM16.2 BD Bioscience
CD94 APC REA113 Miltenyi
CD117 PE-Cy7 104D2 Biolegend
CD161 PerCP-Cy5.5 HP-3G10 Biolegend
CD1c AF700 L161 Biolegend
CD45RA BV510 HI100 Biolegend
CD20 FITC L27 BD Bioscience
CD38 PerCP-Cy5.5 HB7 Biolegend
CD117 PE-Cy7 104D2 Biolegend
CD123 BUV395 7G3 BD Bioscience
CLEC9A PE 8F9 Biolegend
CD3 FITC SK7 BD Bioscience
CD11c BV421 B-Ly6 BD Bioscience
CD14 PE Dazzle HCD14 Biolegend
CD16 FITC NKP15 BD Bioscience
CD19 FITC 4G7 BD Bioscience
CD38 PerCP-Cy5.5 HB-7 Biolegend
CD49f PE-Cy7 GoH3 eBioscience

CD56 FITC NCAM16.2 BD Bioscience
CD90 APC 5E10 Biolegend
CD11b APC ICRF44 Biolegend
CD14 APC-Cy7 HCD14 Biolegend
CD15 BUV395 HI98 BD Bioscience
CD45 V500 HI30 BD Bioscience
CD56 PE NCAM16.2 BD Bioscience
Lineage FITC UCHT1, HCD14, HIB19, 2H7,HCD56 Biolegend
CD11c FITC 3.9 Biolegend
CD38 PE-Cy7 HIT2 Biolegend
CD45RA AF700 HI100 Biolegend
CD34 APC-Cy7 581 Biolegend
CD10 BV421 HI10a BD Bioscience
CD71 BV650 CY1G4 Biolegend
CD45 BV786 2D1 Biolegend
Zombie Aqua - Biolegend
CD45 PE-Cy5 HI30 Biolegend
CD41 FITC HIP8 Biolegend
CD11b APC-Cy7 ICRF44 Biolegend
CD14 PE-Cy7 M5E2 Biolegend
CD15 BV421 MC-480 Biolegend
GYPA PE GA-R2 BD Bioscience
CD56 APC HCD56 Biolegend
CD45 PE-Cy5 HI30 Biolegend
CD19 FITC HIB19 Biolegend
CD19 AF700 HIB19 Biolegend
CD11b APC-Cy7 ICRF44 Biolegend
CD15 BV421 MC-480 Biolegend
CD11c FITC 3.9 Biolegend
CD45RA PE HI100 Biolegend
CD38 PE-Cy7 HIT2 Biolegend
CD90 APC 5E10 BD Bioscience
CD34 APC-Cy7 581 Biolegend
Ki67 FITC B56 BD Bioscience
CD68 KP1 Mouse Biolegend
GYPA R10 Mouse R&D Systems
CD49d polyclonal Rabbit LSBio
VCAM1 1.43C Mouse ThermoFisher
CD1c 2A7C11 170 Er Novus Bio
CD20 L26 141 Pr eBioscience
CD34 Qbend/10 172 Yb Biorad
CD68 KP1 153 Eu Biolegend
CD79A JCB117 + HM47/A9 161 Dy Abcam
CK19 polyclonal 159 Tb Abcam
CSP1/Heppar OCH1E5 167 Er Abcam
GYPA R10 142 Nd R&D Systems
CD34 EP373Y Rabbit Abcam
GYPA EPR8200 Rabbit Abcam

| | |
|---|---|
| Validation | Imaging antibodies were validated against appropriate positive and negative tissue controls. Flow antibodies were validated by the manufacturer. Our flow cytometry and imaging mass cytometry data adhere to the information standards for MIFlowCyt for Flow/Mass cytometry (https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.20623). |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | MS5 - DSMZ, Germany |
| Authentication | None of the cell lines used were authenticated |
| Mycoplasma contamination | Cell lines were not tested for mycoplasma contamination |
| Commonly misidentified lines (See ICLAC register) | *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.* |

# Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | All embryo and fetal tissues were between 4 and 17 post conception weeks. For the liver samples we collected six male and nine female. All tissues were normal as determined by chromosomal assessment by QPCR to exclude the most commonly seen chromosomal abnormalities. |
| Recruitment | Human fetal tissues were obtained from the MRC/Wellcome Trust-funded Human Developmental Biology Resource (HDBR; http://www.hdbr.org) with appropriate written consent and approval from the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee (08/H0906/21+5). HDBR is regulated by the UK Human Tissue Authority (HTA;www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice. Embryos and fetal specimens used for light sheet fluorescence microscopy were obtained with written informed consent from the parents (Gynecology Hospital Jeanne de Flandres, Lille, France) with approval of the local ethic committee (protocol NPFS16-002). Tissues were made available in accordance with the French bylaw (Good practice concerning the conservation, transformation and transportation of human tissue to be used therapeutically, published on December 29, 1998). Permission to utilize human tissues was obtained from the French agency for biomedical research (Agence de la Biome'decine #2016-841, Saint-Denis La Plaine, France). |
| Ethics oversight | Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | All tissues were processed immediately after isolation using the same protocol. Tissue was transferred to a sterile 10mm2 tissue culture dish and cut into <1mm3 segments before being transferred to a 50mL conical tube. Tissue was digested with 1.6mg/mL collagenase type IV (Worthington) in RPMI (Sigma- Aldrich) supplemented with 10%(v/v) heat-inactivated fetal bovine serum (Gibco), 100U/mL penicillin (Sigma-Aldrich), 0.1mg/mL streptomycin (Sigma-Aldrich), and 2mM L-Glutamine (Sigma-Aldrich) for 30 minutes at 37°C with intermittent shaking. Digested tissue was passed through a 100µm filter, and cells collected by centrifugation (500g for 5 minutes at 4°C). Cells were treated with 1X RBC lysis buffer (eBioscience) for 5 minutes at room temperature and washed once with flow buffer (PBS containing 5%(v/v) FBS and 2mM EDTA) prior to counting. Antibody panels were designed to allow enrichment of cell fractions for sequencing and cell types validation. An antibody cocktail was prepared fresh by adding 3µL of each antibody in 50µL Brilliant Stain Buffer (BD) per tissue. Cells (≤10x106) were resuspended in 50-100µL flow buffer and an equal volume of antibody mix was added to cells from each tissue. Cells were stained for 30 minutes on ice, washed with flow buffer and resuspended at 10x106cells/mL. DAPI (Sigma-Aldrich) was added to a final concentration of 3µM immediately prior to sorting. |
| Instrument | Flow sorting was performed on a BD FACSAriaTM Fusion instrument |
| Software | FlowJoV10.4.1 |
| Cell population abundance | Abundance of CD45 positive and negative fractions for droplet single sequencing were determined by cell counting post sort. Purity was checked indirectly by mini bulk RNAseq validation and cytospins for morphology. Additional purity checks for functional experiment (HSC differentiation culture) also included FACS index data and single cell RNA sequencing of sorted cells. |
| Gating strategy | As shown in Extended Data Figure 8, for all flow experiments, cells were gates based on FSC/SSC, live (DAPI negative set based on unstained cells from the sample sample) and single cells (FSC-H/FSC-A). For single cell sequencing, the 'positive' gate was set between the middle of positive and negative staining to the edge of plot, and 'negative' was set to everything else to ensure that all cells were accounted for. For validation experiments (mini bulk RNAseq, cytospins and culture sorts), gates were set over the bulk of the positive staining excluding the edges of staining. Our flow cytometry data adhere to the information standards for Flow cytometry (https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.20623). |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

https://doi.org/10.1038/s41586-019-1653-x

# Habenular TCF7L2 links nicotine addiction to diabetes

Alexander Duncan[1,2,9], Mary P. Heyer[1,9], Masago Ishikawa[1,9], Stephanie P. B. Caligiuri[1,9], Xin-an Liu[1,8], Zuxin Chen[1,8], Maria Vittoria Micioni Di Bonaventura[1], Karim S. Elayouby[1], Jessica L. Ables[1], William M. Howe[1], Purva Bali[1], Clementine Fillinger[1], Maya Williams[1], Richard M. O'Connor[1], Zichen Wang[3], Qun Lu[4], Theodore M. Kamenecka[4], Avi Ma'ayan[3], Heidi C. O'Neill[5], Ines Ibanez-Tallon[6], Aron M. Geurts[7] & Paul J. Kenny[1]*

**Diabetes is far more prevalent in smokers than non-smokers, but the underlying mechanisms of vulnerability are unknown. Here we show that the diabetes-associated gene *Tcf7l2* is densely expressed in the medial habenula (mHb) region of the rodent brain, where it regulates the function of nicotinic acetylcholine receptors. Inhibition of TCF7L2 signalling in the mHb increases nicotine intake in mice and rats. Nicotine increases levels of blood glucose by TCF7L2-dependent stimulation of the mHb. Virus-tracing experiments identify a polysynaptic connection from the mHb to the pancreas, and wild-type rats with a history of nicotine consumption show increased circulating levels of glucagon and insulin, and diabetes-like dysregulation of blood glucose homeostasis. By contrast, mutant *Tcf7l2* rats are resistant to these actions of nicotine. Our findings suggest that TCF7L2 regulates the stimulatory actions of nicotine on a habenula–pancreas axis that links the addictive properties of nicotine to its diabetes-promoting actions.**

Nicotine is the major reinforcing component of tobacco responsible for addiction in cigarette smokers[1]. The stimulatory action of nicotine on dopamine neurons in the ventral tegmental area is considered necessary and sufficient for the rewarding effects of the drug that motivate smoking behaviour[2]. Nicotine also activates cholinergic neurons in the mHb that project to the interpeduncular nucleus (IPn), which elicits noxious responses to nicotine[3]. The mechanisms that regulate the stimulatory actions of nicotine on mHb neurons remain poorly understood[4]. This is important because sensitivity to the noxious effects of nicotine has a crucial role in determining the likelihood of progressing from initial to habitual tobacco use and, once the habit is established, the amounts of tobacco that are consumed[4]. Nicotine was recently shown[5] to activate neurons in the hindbrain that synthesize GLP-1. These hindbrain neurons project to the IPn, where locally released GLP-1 enhances mHb-derived excitatory transmission[5]. This action abolishes nicotine reward and promotes nicotine avoidance behaviours[5]. TCF7L2 is considered a core component of the GLP-1 signalling cascade in pancreatic β cells and other tissues[6], but the function of TCF7L2 in the brain is unclear. TCF7L2 is expressed in all major tissues involved in glycaemic control, including the pancreas and liver[7], and allelic variation in *TCF7L2* is among the most strongly associated and best-replicated genetic risk factors for type 2 diabetes[8,9]. Notably, nicotine contained in cigarettes can stimulate increases in blood glucose levels[10,11], and habitual tobacco smoking markedly increases the risk of type 2 diabetes[12,13]. How nicotine increases blood glucose or the relevance of this action to either the persistence of the smoking habit or the pathophysiology of diabetes in smokers is unknown. Here, we investigated the role of TCF7L2 in regulating the motivational properties of nicotine and explored the link between habenular TCF7L2 activity and the diabetes-promoting actions of nicotine.

## Habenular TCF7L2 regulates nicotine intake

Cholinergic neurons in the mHb co-release glutamate, are the major source of excitatory transmission in the IPn, and have a key role in controlling nicotine intake[5,14]. Using translating ribosome affinity purification (TRAP) data collected from the mHb of *Chat*$^{DW167}$ mice, in which enhanced green fluorescent protein (eGFP)-tagged L10a ribosomal subunit is expressed in cholinergic neurons[15], we found that *Tcf7l2* transcripts were highly enriched in mHb cholinergic neurons (Fig. 1a, b), and were 4–6-fold higher in the mHb than in the striatum, frontal cortex or hippocampus[16,17] (Fig. 1c). TCF7L2 protein was also densely expressed in the mHb compared with surrounding brain regions in ChAT-tdTom mice (Fig. 1d), in which the fluorescent reporter tdTomato is expressed in mHb cholinergic neurons. TCF7L2 immunofluorescence colocalized with tdTomato-positive and tdTomato-negative cells in the mHb (Fig. 1d), which suggests that TCF7L2 is expressed by both cholinergic and non-cholinergic cells. Robust activity of β-galactosidase was detected in the mHb of BAT-GAL mice, in which expression of β-galactosidase is controlled by TCF7L2 (Fig. 1e). Hence, TCF7L2 is robustly expressed and functionally active in habenular cholinergic neurons that regulate nicotine intake[5,14].

Next, we investigated the role of TCF7L2 in regulating the motivational properties of nicotine. Mice with a null mutation in *Tcf7l2* (*Tcf7l2*$^{-/-}$) die during the early postnatal period[18,19]. Therefore, we generated a line of *Tcf7l2* mutant rats using zinc-finger nucleases (SS-*Tcf7l2*$^{em1Mcwi}$; RGD ID: 5509993)[20]. Specifically, we deleted the β-catenin binding domain of the rat *Tcf7l2* gene (Extended Data Fig. 1, Methods), a site crucial for stimulating its transcriptional activity[21] (Fig. 1f). Mutant *Tcf7l2* rats survived to adulthood, showed no obvious deleterious health effects, and did not have broad structural or functional abnormalities in the mHb–IPn circuit (Extended Data Fig. 1). Mutant *Tcf7l2* rats responded far more vigorously than wild-type

---

[1] Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [2] Skaggs Graduate School of Chemical and Biological Sciences, Scripps Research, Jupiter, FL, USA. [3] Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4] Department of Molecular Medicine, The Scripps Research Institute, Jupiter, FL, USA. [5] Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA. [6] The Laboratory of Molecular Biology, The Rockefeller University, New York, NY, USA. [7] Department of Physiology, Medical College of Wisconsin, Milwaukee, WI, USA. [8] Present address: Brain Cognition and Brain Disease Institute (BCBDI), Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen, China. [9] These authors contributed equally: Alexander Duncan, Mary P. Heyer, Masago Ishikawa, Stephanie P. B. Caligiuri. *e-mail: paul.kenny@mssm.edu

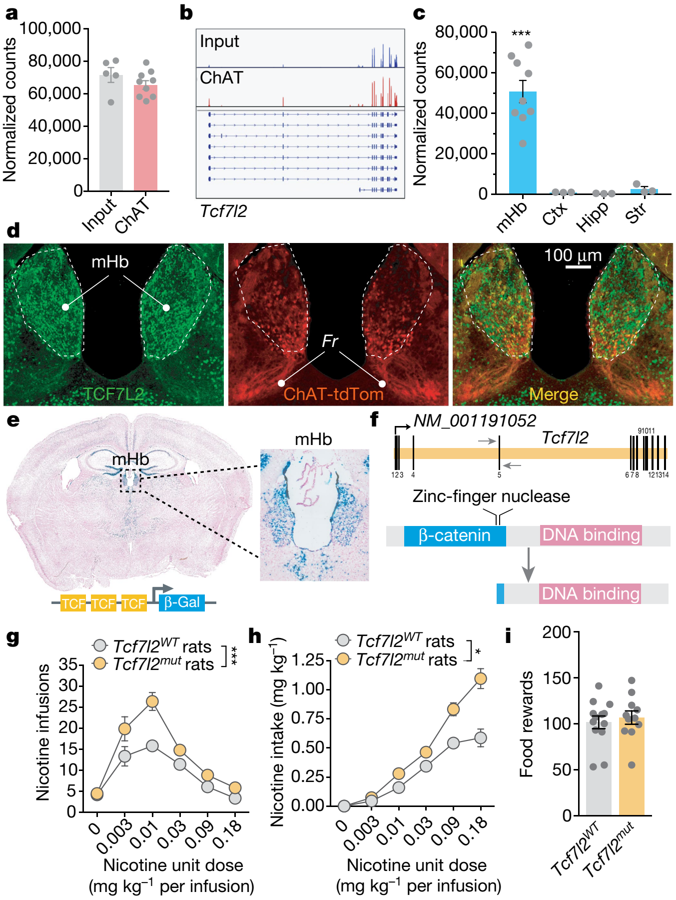


**Fig. 1 | TCF7L2 is enriched in the medial habenula and regulates nicotine intake.** **a**, Bacterial artificial chromosome-based TRAP (bacTRAP) from *Chat*$^{DW167}$ mice showed that TCF7L2 is highly expressed in habenular cholinergic cells of mice compared with total habenular input (pooled samples from $n = 9$ *Chat*$^{DW167}$ mice for immunoprecipitation and $n = 5$ *Chat*$^{DW167}$ mice for input). **b**, Example of RNA-seq reads from habenula of *Chat*$^{DW167}$ TRAP mice aligned to the *Tcf7l2* gene; observation replicated in subsequent TRAP experiments. **c**, RNA-seq showed that TCF7L2 reads are higher in the mHb (from $n = 9$ mice) than in the cortex (ctx; $n = 3$ mice), hippocampus (hipp; $n = 3$ mice) or striatum (str; $n = 3$ mice) of mice ($F_{3,14} = 23.6$, $P < 0.001$; one-way analysis of variance (ANOVA); $***P < 0.001$ compared with each of the other brain regions, Bonferroni's multiple comparisons test). **d**, Immunofluorescence detection of TCF7L2 (green) in the mHb of ChAT-tdTom (red) reporter mice. **e**, β-Gal activity in the mHb of BAT-GAL reporter mice (seen independently in $n = 3$ mice). **f**, Strategy for deleting the β-catenin-binding domain of the *Tcf7l2* gene using zinc-finger nucleases in rats. **g**, Responding for nicotine was increased in mutant (*Tcf7l2*$^{mut}$) rats ($n = 30$) compared with wild-type (*Tcf7l2*$^{WT}$) rats ($n = 22$) ($F_{1,236} = 32.75$, $***P < 0.0001$, main effect of 'genotype' in two-way ANOVA). **h**, Total nicotine intake was increased in mutant rats compared with wild-type *Tcf7l2* rats ($F_{1,193} = 6.72$; $*P = 0.0102$, main effect of 'genotype' in two-way ANOVA). **i**, Responding for food rewards was similar in mutant and wild-type *Tcf7l2* rats ($P = 0.61$, unpaired two-tailed *t*-test). Data are mean ± s.e.m.

*Tcf7l2* rats to intravenous nicotine infusions (0.03–0.18 mg kg$^{-1}$ per infusion) (Fig. 1g). Wild-type *Tcf7l2* rats titrated their responses to earn a maximum of approximately 0.6 mg kg$^{-1}$ during the 1 h sessions (Fig. 1h), whereas mutant *Tcf7l2* rats showed far less restraint over their nicotine intake (Fig. 1h), which progressively increased across sessions (Extended Data Fig. 1). Mutant and wild-type *Tcf7l2* rats responded at similar rates for food reinforcers (45-mg pellets) (Fig. 1i), and behaved similarly in an open-field arena after acute nicotine injection (0.4 mg kg$^{-1}$) (Extended Data Fig. 1). These findings suggest that TCF7L2 deficiency increases nicotine intake and this effect is not secondary to alterations in behavioural or motor performance.

Next, we used the CRISPR–Cas9 system to investigate the role of TCF7L2 in the mHb in regulating nicotine intake. We delivered an adeno-associated virus (AAV) to express a short guide RNA (sgRNA) that cleaves mouse *Tcf7l2* DNA (AAV-sgRNA-Tcf7l2) (Extended Data Fig. 2) or a control virus (AAV-sgRNA-eGFP), along with an AAV to express Cre recombinase (AAV2-hSYN1-iCre), into the mHb of *Rosa26*$^{LSL\text{-}spCas9\text{-}eGFP}$ mice, which express Cas9 in a Cre-dependent manner (Fig. 2c, d, Extended Data Fig. 2). *Rosa26*$^{LSL\text{-}spCas9\text{-}eGFP}$ mice treated with AAV-sgRNA-Tcf7l2 self-administered far greater quantities of nicotine than those treated with AAV-sgRNA-eGFP control (Fig. 2e, f, Extended Data Fig. 2), whereas their responses for food rewards (25-mg pellets) were similar (Fig. 2g). Short interfering RNA (siRNA)-mediated knockdown of *Tcf7l2* transcripts in the mHb also increased nicotine intake in rats (Fig. 2h–j). In addition to being a core component of the GLP-1 signalling cascade, TCF7L2 is also activated by Wnt glycoproteins[22] and insulin[23]. Infusion of the GLP-1 receptor agonist exendin-4 (EX-4; 12.5–100 ng) into the mHb reduced nicotine intake (Extended Data Fig. 2). By contrast, mHb infusion of DKK1 (100 ng), a secreted endogenous inhibitor of Wnt signalling, XAV939 (12.5 ng), a small molecule inhibitor of Wnt signalling[24], or insulin (30 nM) did not alter nicotine intake (Extended Data Fig. 2). These data establish a key role for habenular TCF7L2 in controlling nicotine intake and suggest that GLP-1 is likely to regulate habenular TCF7L2 activity.

## TCF7L2 regulates nAChRs in habenular neurons

Next, we investigated the mechanisms by which TCF7L2 acts in the mHb to control nicotine intake. Nicotine increased the frequency, but not the amplitude, of spontaneous excitatory post-synaptic currents (sEPSCs) in IPn neurons when applied to the mHb in slices from wild-type *Tcf7l2* rats[25] (Fig. 2m–o). This effect was almost completely absent in mutant *Tcf7l2* rats (Fig. 2m, n), which suggests that TCF7L2 regulates the ability of nicotine to activate the mHb–IPn circuit. Nicotine had no effects on the transcriptional activity of TCF7L2, or the activation status of its transactivator β-catenin, in rat PC12 cells (Extended Data Fig. 3). Similarly, nicotine (0.125–1.0 mg kg$^{-1}$) did not alter TCF7L2 activity in the mHb of BAT-GAL mice, and self-administration of nicotine (0.18 mg kg$^{-1}$ per infusion) also did not alter habenular TCF7L2 levels in rats (Extended Data Fig. 3, Supplementary Fig. 1). Hence, nicotine is unlikely to stimulate excitatory transmission in the IPn by a mechanism that involves TCF7L2 activation. Increases in intracellular calcium concentration ($[Ca^{2+}]_i$) evoked by nicotine in human HEK293T cells that stably express α4β2α5 nicotinic acetylcholine receptors (nAChRs), which are considered one of the major subtypes of nAChRs in the mHb–IPn circuit[26], were attenuated by siRNA-mediated knockdown of TCF7L2 (Extended Data Fig. 3). Furthermore, acetylcholine-evoked increases in radiolabelled rubidium ($^{86}Rb^+$) efflux, considered a measure of presynaptic nAChR function[3], were attenuated in synaptosomes from the IPn of mutant compared with wild-type *Tcf7l2* rats (Extended Data Fig. 3). This suggests that TCF7L2 regulates the function of habenular nAChRs. To investigate this possibility directly, we pharmacologically isolated nAChR currents in mHb neurons from wild-type and mutant *Tcf7l2* rats. The magnitude of nAChR currents evoked by low frequency pressure-application of nicotine (30 μM; pulsed at 0.1 Hz) (Fig. 3a), was similar in mHb neurons from wild-type and mutant *Tcf7l2* rats (Fig. 3b, c, Extended Data Fig. 3). Habenular nAChRs desensitized with the same temporal dynamics in wild-type and mutant *Tcf7l2* rats when the frequency of nicotine pulses was increased (0.1–1.0 Hz) (Fig. 3b, Extended Data Fig. 3). Notably, nAChR currents rapidly recovered in mHb neurons from wild-type *Tcf7l2* rats when the frequency of nicotine application was decreased (1.0–0.1 Hz), whereas nAChR currents did not recover in mutant *Tcf7l2* rats (Fig. 3b, c). These findings reveal a crucial role for TCF7L2 in regulating the function of habenular nAChRs by regulating their capacity to recover from nicotine-induced desensitization.

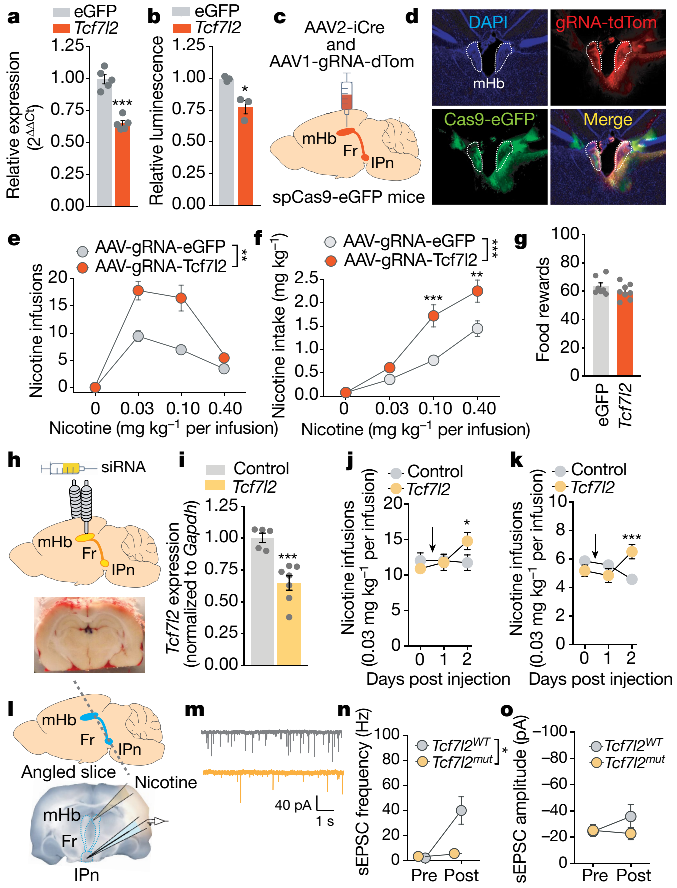


**Fig. 2 | TCF7L2 regulates habenular sensitivity to nicotine. a**, CRISPR-mediated cleavage of *Tcf7l2* lowers *Tcf7l2* mRNA in mouse neuroblastoma N2a cells (from three independent experiments) (***$P < 0.001$, unpaired two-tailed *t*-test). **b**, *Tcf7l2* cleavage decreases TCF7L2 transcriptional activity (from three independent experiments) (*$P = 0.014$, unpaired two-tailed *t*-test). **c**, Graphical representation showing delivery of Cre-expressing and sgRNA-expressing viruses to the mHb. Fr, fasciculus retroflexus. **d**, DAPI-counterstained brain slice from a $Rosa^{LSL\text{-}spCas9\text{-}eGFP}$ mouse showing GFP- and tdTomato-labelled cells in the mHb. Representative result from $n = 3$ mice. **e**, Responding for nicotine was increased in $Rosa^{LSL\text{-}spCas9\text{-}eGFP}$ mice treated with AAV-sgRNA-Tcf7l2 ($n = 8$) compared with those treated with AAV-sgRNA-eGFP ($n = 7$) ($F_{1,39} = 34.2$; ***$P < 0.0001$, main effect in two-way ANOVA). **f**, Total nicotine intake was increased in $Rosa^{LSL\text{-}spCas9\text{-}eGFP}$ mice treated with AAV-sgRNA-Tcf7l2 compared with those treated with AAV-sgRNA-eGFP ($F_{1,13} = 16.98$; ***$P < 0.005$, main effect in two-way ANOVA). **g**, Responding for food rewards was similar in $Rosa^{LSL\text{-}spCas9\text{-}eGFP}$ mice treated with AAV-sgRNA-Tcf7l2 ($n = 8$) and mice treated with AAV-sgRNA-eGFP ($n = 7$). **h**, Graphical representation of rat brain showing cannula above the mHb (top) and coronal brain slice showing accurate targeting of dye into mHb. **i**, Confirmation of siRNA-mediated TCF7L2 knockdown using qPCR ($n = 5$ biologically independent control rats; $n = 6$ biologically independent siRNA rats) (***$P = 0.0007$, unpaired two-tailed *t*-test). **j**, TCF7L2 knockdown in the mHb increased nicotine intake (0.03 mg kg$^{-1}$ per infusion; $n = 8$ biologically independent rats) ($F_{2,26} = 5.03$; *$P = 0.0142$; interaction effect in two-way ANOVA). **k**, siRNA-mediated knockdown of TCF7L2 in the mHb increased nicotine intake (0.12 mg kg$^{-1}$ per infusion; $n = 7$ biologically independent rats) ($F_{2,22} = 7.52$; *$P = 0.0007$; interaction effect in two-way ANOVA). **l**, Graphical representation of angled brain slice containing the mHb, fasciculus retroflexus and IPn (top) and representative brain slice showing position of pipettes for nicotine delivery and recording (bottom) **m**, Representative traces of sEPSCs recorded in IPn neurons from wild-type and mutant *Tcf7l2* rats after delivery of nicotine to mHb neurons. **n**, Nicotine-induced increases in sEPSCs frequency were lower in the IPn of mutant rats than wild-type rats ($F_{1,16} = 8.08$; *$P = 0.0118$, interaction effect in two-way ANOVA). **o**, Nicotine did not alter the amplitude of sEPSCs in IPn of wild-type or mutant *Tcf7l2* rats. Data are mean ± s.e.m.

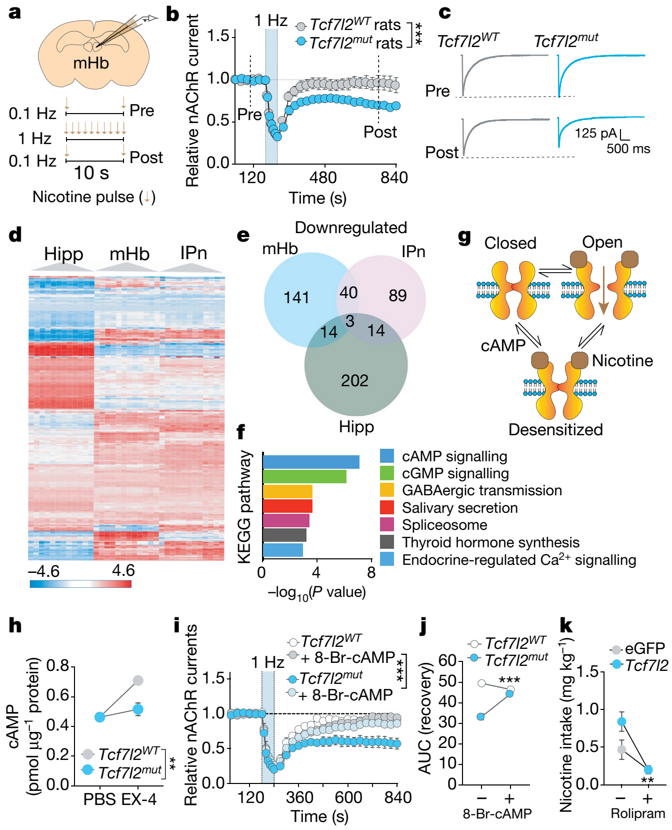


**Fig. 3 | TCF7L2 regulates habenular nAChR function. a**, Representation of recording sites in the mHb (top) and protocol for nicotine delivery (bottom). Image adapted from the Allen Brain Reference Atlas. **b**, Relative nAChR-mediated currents in response to low-frequency (0.1 Hz; pre), high-frequency (1 Hz), then low-frequency (0.1 Hz; post) pulses of nicotine (30 µM) in mHb neurons from wild-type (10 cells from 4 rats) and mutant (11 cells from 4 rats) *Tcf7l2* rats ($F_{31,558} = 5.42$; ***$P < 0.001$, interaction effect in two-way ANOVA). Dotted lines identify time points at which representative traces shown in **c** were collected. **c**, Representative nAChR current traces in mHb neurons from wild-type and mutant *Tcf7l2* rats evoked by nicotine before (Pre) and after (Post) high-frequency nicotine pulses to desensitize nAChRs. **d**, Heat map of RNA-seq expression data from the hippocampus, mHb and IPn of wild-type and mutant *Tcf7l2* rats ($n = 9$ rats per genotype). Displayed are 600 most differentially expressed genes clustered according to brain region. Data are log-transformed and *z*-scored (shown in scale). **e**, Venn diagram of differentially downregulated genes in the hippocampus, mHb and IPn. **f**, KEGG analysis of differentially downregulated genes suggests that cAMP signalling is likely to be perturbed in the mHb of mutant *Tcf7l2* rats (9 rats per genotype); *P* values determined by Fisher exact test. **g**, Representation of nAChR in the closed (inactive) conformation, nicotine-induced stabilization of the open (active) confirmation, and entry into a desensitized state from which cAMP facilitates recovery. **h**, cAMP content of mHb tissues from wild-type and mutant *Tcf7l2* rats treated with PBS or EX-4 (100 nM) (3 rats per replicate; 9 rats per genotype) ($F_{1,8} = 13.08$; **$P = 0.0068$, interaction effect in two-way ANOVA). **i**, Relative nAChR-mediated currents in response to low-, high- and then low-frequency pulses of nicotine in neurons from wild-type (6 cells, 4 rats), mutant (6 cells, 3 rats), wild-type + 8-Br-cAMP (200 µM) (6 cells, 3 rats) and mutant + 8-Br-cAMP (6 cells, 4 rats) rats ($F_{31,640} = 2.42$; ***$P < 0.0001$; interaction effect in three-way ANOVA). **j**, Area under the curve (AUC) analysis of nAChR recovery from desensitization from the time of maximal desensitization (240 s) to the end of the recording period (840 s) in mHb neurons from wild-type and mutant *Tcf7l2* rats; ($F_{1,20} = 44.1$; ***$P < 0.0001$; interaction effect two-way ANOVA). AUC analysis from 6 wild-type cells (from 3 rats) and 6 mutant cells (from 3 rats). **k**, Rolipram decreased nicotine intake in $Rosa^{LSL\text{-}spCas9\text{-}eGFP}$ mice that received intra-mHb injections of AAV-sgRNA-Tcf7l2 ($F_{1,9} = 20.9$; ***$P < 0.0001$; interaction effect in two-way ANOVA; $n = 5$ AAV-sgRNA-Tcf7l2 mice, $n = 6$ AAV-sgRNA-eGFP mice). Data are mean ± s.e.m.

## TCF7L2 regulates cAMP signalling in the habenula

Because TCF7L2 is a transcription factor[27], we used RNA sequencing (RNA-seq) analysis to identify differentially expressed genes in the mHb of mutant *Tcf7l2* rats that may explain its actions on nAChR function. For comparison, RNA from the IPn and hippocampus was also sequenced (Fig. 3d). This analysis identified approximately 195 genes that were differentially downregulated in the mHb of mutant *Tcf7l2* rats compared with wild-type *Tcf7l2* rats, of which 141 were specifically downregulated in the mHb compared with the IPn or hippocampus (Fig. 3e). This pattern of differential expression was confirmed for representative genes by quantitative PCR (qPCR) analysis (Extended Data Fig. 4). Notably, nAChR subunit transcript levels were similar in the mHb of mutant relative to wild-type *Tcf7l2* rats (Extended Data Fig. 4), which suggests that TCF7L2 regulates the function but not the expression of habenular nAChRs. siRNA-mediated knockdown of five of these differentially downregulated genes (*PAFAH1B1*, *NDFIP1*, *ARHGAP5*, *HNRNPU* and *AKAP9*) did not alter $\alpha4\beta2\alpha5$ nAChR function in human HEK293T cells (Extended Data Fig. 4). However, Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of the differentially downregulated genes predicted that habenular cAMP signalling is likely to be perturbed in mutant *Tcf7l2* rats ($P = 9.78 \times 10^{-9}$) (Fig. 3f). Consistent with this prediction, a dominant-negative *Tcf7l2* mutant reduced the activity of a cAMP luciferase reporter in rat PC12 cells (Extended Data Fig. 5). Similarly, dominant-negative *Tcf7l2* reduced baseline and EX-4-evoked increases in cAMP luciferase reporter activity in INS-1 cells—an immortalized rat pancreatic β cell line that constitutively expresses GLP-1 receptors[28] (Extended Data Fig. 5). EX-4 increased the activity of a lentivirus-expressed cAMP luciferase reporter in the mHb of wild-type but not mutant *Tcf7l2* rats (Extended Data Fig. 5). Furthermore, EX-4-evoked increases in cAMP levels in mHb tissues, measured by enzyme-linked immunosorbent assay (ELISA), were greatly attenuated in mHb tissue from mutant *Tcf7l2* rats compared with wild-type *Tcf7l2* rats (Fig. 3h). The cAMP analogue 8-bromo-cAMP (8-Br-cAMP) facilitated the recovery of $\alpha4\beta2\alpha5$ nAChRs stably expressed in HEK293T cells from nicotine-induced desensitization (Extended Data Fig. 5), consistent with cAMP-dependent kinases regulating this process[29–31]. Moreover, 8-Br-cAMP also rescued the failure of nAChRs to recover from nicotine-induced desensitization in mHb neurons from mutant *Tcf7l2* rats (Fig. 3i, j). Finally, the phosphodiesterase inhibitor rolipram (1 mg $\mathrm{kg}^{-1}$), which increases cAMP levels in the brains of mice[32], rescued the otherwise increased nicotine intake in *Rosa26*$^{LSL\text{-}spCas9\text{-}eGFP}$ mice in which habenular *Tcf7l2* was cleaved (Fig. 3k). Together, these findings identify a crucial role for TCF7L2 in regulating the function of habenular nAChRs, and hence nicotine intake, through a mechanism involving control of local cAMP signalling dynamics.

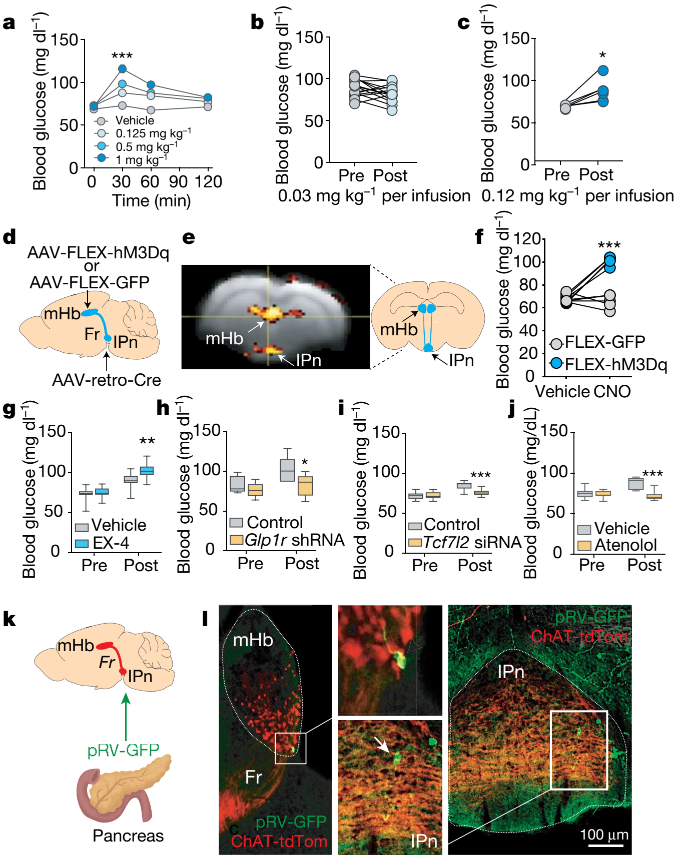


**Fig. 4 | Habenular TCF7L2 regulates hyperglycaemic response to nicotine.** **a**, Nicotine increased fasting blood glucose levels in rats ($F_{9,66} = 13$; $^{***}P < 0.0001$; interaction effect in two-way ANOVA; $n = 8$). **b**, Self-administration of the standard dose of nicotine (0.03 mg $\mathrm{kg}^{-1}$ per infusion) did not alter blood glucose levels in rats ($n = 15$). **c**, Self-administration of a higher nicotine dose (0.12 mg $\mathrm{kg}^{-1}$ per infusion) increased blood glucose in rats ($n = 6$). $^{**}P = 0.0092$, two-tailed paired $t$-test. **d**, Graphical representation of rat brain showing strategy to stimulate the mHb–IPn circuit chemogenetically. Retro-AAV-Cre was delivered into the IPn, and Cre-dependent FLEX-hM3Dq ($n = 5$) or FLEX-GFP ($n = 5$) was delivered into the mHb. **e**, Clozapine-*N*-oxide (CNO; 3 mg $\mathrm{kg}^{-1}$) increased the activity of the mHb–IPn circuit of rats that express FLEX-hM3Dq but not in those that express FLEX-GFP, reflected by changes in the BOLD signal measured by functional magnetic resonance imaging. Experiment was performed on a single occasion. Coronal brain image adapted from the Allen Brain Reference Atlas. **f**, CNO injection increased fasting blood glucose levels in rats that express FLEX-hM3Dq ($n = 5$ rats) compared with those that express FLEX-GFP ($n = 5$ rats) ($F_{1,8} = 85.2$; $^{***}P < 0.0001$; interaction effect in two-way ANOVA). **g**, Intra-mHb infusion of EX-4 (100 ng) enhanced the hyperglycaemic response to self-administered nicotine in rats ($n = 10$) ($F_{1,21} = 8.39$; $^{**}P = 0.0086$; interaction effect in two-way ANOVA). **h**, Knockdown of *Glp1r* transcripts in the mHb abolished the hyperglycaemic response to self-administered nicotine infusions in rats ($n = 6$ per virus); $F_{1,10} = 5.15$; $^{*}P = 0.0466$; interaction effect in two-way ANOVA. **i**, TCF7L2 knockdown in the mHb abolished the hyperglycaemic response to self-administered nicotine in rats ($n = 10$) ($F_{1,18} = 18.6$; $^{***}P < 0.001$; interaction effect in two-way ANOVA). **j**, Atenolol abolished the hyperglycaemic response to self-administered nicotine in rats ($n = 7$) ($F_{1,12} = 19.8$; $^{***}P < 0.001$; interaction effect in two-way ANOVA). Box plots in **g**–**j** show show minimum–maximum range. **k**, Graphical representation of strategy to trace polysynaptic inputs from the mHb–IPn circuit to the pancreas in ChAT-tdTom reporter mice. **l**, Image of a pRV-GFP-labelled cholinergic cell in the ventral region of the mHb, and pRV-GFP-labelled IPn neuron in close apposition to cholinergic fibres (red) from the mHb (inserts show magnified images). Representative result from two independent experiments.

## Habenular TCF7L2 sets glycaemic responses to nicotine

In addition to downregulated genes, we detected 161 differentially upregulated genes in the mHb of mutant relative to wild-type *Tcf7l2* rats, 85 of which were specifically upregulated in the mHb compared with the IPn or hippocampus (Extended Data Fig. 6). KEGG analysis of these upregulated genes suggested that five of the six top pathways predicted to be perturbed are involved in energy homeostasis, particularly regulation of glucose metabolism and blood glucose levels (Extended Data Fig. 6). Considering that nicotine can increase blood glucose in smokers[10,11], smoking is a major risk factor for type 2 diabetes[12,13], and *TCF7L2* alleles are strongly associated with type 2 diabetes[8,9], we proposed that TCF7L2 signalling in the mHb may regulate the actions of nicotine on blood glucose. Nicotine (0.25–1.00 mg $\mathrm{kg}^{-1}$) dose-dependently increased blood glucose levels in rats (Fig. 4a). By contrast, systemic injection of oxycodone (2.5 mg $\mathrm{kg}^{-1}$) or cocaine (20 mg $\mathrm{kg}^{-1}$) had no effect (Extended Data Fig. 6). Blood glucose was unaltered in rats by self-administration of the standard nicotine dose used in most experiments (0.03 mg $\mathrm{kg}^{-1}$ per infusion) (Fig. 4b). However, a higher unit dose (0.12 mg $\mathrm{kg}^{-1}$ per infusion), known to stimulate activity in the mHb–IPn circuit[3], increased blood glucose (Fig. 4c). Chemogenetic stimulation of the mHb–IPn circuit similarly increased blood glucose in rats, mimicking the effects of nicotine (Fig. 4d–f, Extended Data Fig. 7). Chemogenetic stimulation of the IPn of *Chrna5-cre* transgenic mice, which express Cre-recombinase from the nAChR gene *Chrna5*, also increased blood glucose levels (Extended Data Fig. 6). Infusion of EX-4 (100 ng) into the mHb had no effects on blood glucose in rats that self-administered saline (Fig. 4g) but enhanced the hyperglycaemic response to nicotine (0.12 mg $kg^{-1}$ per infusion) (Fig. 4g). Conversely, short hairpin RNA (shRNA)-mediated knockdown of GLP-1 receptors in the mHb reduced local TCF7L2 expression and abolished the stimulatory effects of nicotine on blood glucose (Fig. 4h, Extended Data Fig. 6). siRNA-mediated knockdown of TCF7L2 in the mHb also reduced the hyperglycaemic actions of nicotine (Fig. 4i). Similarly, the $\beta_1$-adrenergic receptor ($\beta_1$AR) antagonist atenolol (10 mg $kg^{-1}$), and the $\beta_2$AR antagonist ICI118,551 (2 mg $kg^{-1}$) blocked the hyperglycaemic response to nicotine injection (Extended Data Fig. 6). Atenolol, which does not cross the blood–brain barrier[33,34], also blocked the hyperglycaemic response to nicotine self-administration (Fig. 4j) or chemogenetic stimulation of the mHb–IPn circuit (Extended Data Fig. 6). Mice treated with a hyperglycaemic dose of nicotine (0.5 mg $kg^{-1}$) had depleted glucagon levels in pancreatic islets (Extended Data Fig. 8). Finally, injection into the pancreas—but not the liver—of pseudorabies virus expressing green fluorescent protein (pRV-GFP), which travels in a retrograde fashion from sites of infection via synaptically connected neurons (Fig. 4k), resulted in GFP-expressing tdTomato-positive cholinergic cells in the mHb and GFP-positive cells in the IPn that were in close apposition to tdTomato-positive fibres from the mHb in these mice (Fig. 4i, Extended Data Fig. 8). Together, these findings suggest that nicotine increases blood glucose levels in a GLP-1- or TCF7L2-dependent manner, by a mechanism that involves habenular recruitment of sympathetic nervous system transmission and release of glucagon from the pancreas.

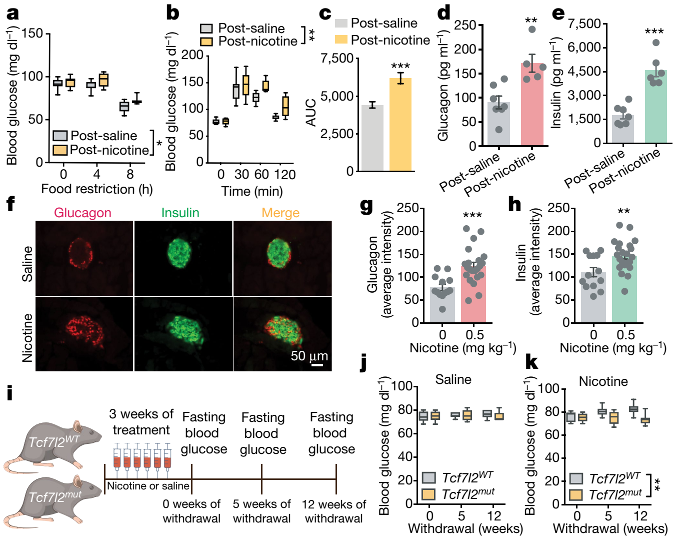


**Fig. 5 | TCF7L2 regulates the diabetes-promoting actions of nicotine.** **a**, Blood glucose levels were assessed over 8 h of food restriction in rats with a history of self-administering saline ($n = 8$) or nicotine ($n = 7$) ($F_{2,26} = 71.1$; $P < 0.0001$, main effect of 'time'; $F_{1,13} = 8.6$; $*P < 0.05$, main effect of 'nicotine' in two-way ANOVA). **b**, Glucose clearance was impaired in nicotine-treated ($n = 6$) compared with saline-treated ($n = 8$) rats in an oral glucose tolerance test. Data are presented as change in blood glucose from time = 0 ($F_{1,12} = 6.2$; $*P = 0.0284$, main effect in two-way ANOVA. **c**, AUC analysis of the time-course of glucose clearance in nicotine-treated ($n = 6$) and saline-treated ($n = 8$) rats in an oral glucose tolerance test ($***P = 0.0007$, unpaired two-tailed *t*-test). **d**, Circulating glucagon levels were increased in nicotine-treated rats ($n = 6$) compared with saline-treated rats ($n = 5$). Data from biologically independent animals ($**P = 0.0042$, unpaired two-tailed *t*-test). **e**, Circulating insulin levels were increased in nicotine-treated rats ($n = 6$) compared with saline-treated rats ($n = 6$). Data from biologically independent animals ($***P < 0.001$, unpaired two-tailed *t*-test). **f**, Immunostaining for glucagon (left), insulin (middle) and their overlap (right) in mice treated chronically with saline (top; $n = 3$) or nicotine (0.5 mg $kg^{-1}$; bottom; $n = 3$). Experiment was performed on a single occasion. **g**, Quantification of glucagon intensity in pancreatic islets from mice treated with saline (from 13 islets images in $n = 3$ mice) or nicotine (from 22 islets images in $n = 3$ mice) ($***P = 0.0006$, unpaired two-trailed *t*-test). **h**, Quantification of insulin intensity in pancreatic islets from mice treated with saline (from 13 islets images in $n = 3$ mice) or nicotine (from 22 islets images in $n = 3$ mice) ($**P < 0.006$, unpaired two-trailed *t*-test). **i**, Graphical representation of experiment designed to test the effects of withdrawal from chronic nicotine injections on blood glucose in wild-type and mutant *Tcf7l2* rats. **j**, Fasting blood glucose levels were similar in wild-type ($n = 6$) and mutant ($n = 6$) *Tcf7l2* rats that were treated chronically with saline. **k**, Fasting blood glucose levels were increased in wild-type rats ($n = 6$) compared with mutant rats ($n = 8$) after chronic nicotine treatment (1 mg $kg^{-1}$) ($F_{1,12} = 10.4$; $**P = 0.0073$, main effect in two-way ANOVA). Box plots show minimum–maximum range. Data are mean ± s.e.m.

## Glucose regulates nAChR function in the habenula

Next, we investigated the behavioural and physiological importance of the hyperglycaemic actions of nicotine. Neither oral nor intravenous infusion of glucose or glucagon had any effect on nicotine self-administration in rats (Extended Data Fig. 9). Similarly, a dose of atenolol that completely blocked the hyperglycaemic actions of nicotine did not alter nicotine intake in rats (Extended Data Fig. 9). This suggests that obtaining the stimulatory effects of nicotine on blood glucose is unlikely to contribute to the motivational properties of the drug. Notably, we identified 1,160 transcripts for which translation was altered in mHb cholinergic neurons of $Chat^{DW167}$ TRAP mice after 6 weeks of daily sucrose consumption (Extended Data Fig. 9). This suggests that increased circulating glucose levels can modulate mHb function. Increasing the glucose concentration in the extracellular solution (12.5–30.0 mM) had no effects on the spike frequency of mHb neurons but decreased the magnitude of nicotine-evoked habenular nAChR currents (Extended Data Fig. 9). This suggests that nicotine-induced increases in blood glucose may 'feedback' onto mHb neurons to inhibit local nAChR function and promote the development of habitual tobacco smoking. Glucose regulation of habenular nAChR function could also explain the high rates of tobacco use in individuals with type 2 diabetes and their greater difficulty quitting the habit than non-diabetic subjects[35]. Finally, we investigated the effects of repeated exposure to the hyperglycaemic actions of nicotine on blood glucose homeostasis. Fasting blood glucose levels were reduced in the rats with a history of nicotine self-administration (0.12 mg $kg^{-1}$ per infusion; 21 daily sessions) compared with rats that self-administered saline when measured 24 h after their final session (Extended Data Fig. 9). This hypoglycaemic state during early nicotine withdrawal probably contributes to the well-known carbohydrate craving experienced by tobacco smokers during the initial stages of an attempt to quit smoking[36]. By contrast, the nicotine-experienced rats showed increased fasting blood glucose levels and deficits in glucose clearance compared with saline-experienced rats when measured three and six weeks after their final session (Fig. 5a–c, Extended Data Fig. 9). Circulating levels of glucagon and insulin were also increased in the nicotine-experienced rats (Fig. 5d, e). Similarly, pancreatic glucagon and insulin content were increased in mice after more than one month of withdrawal from chronic nicotine treatment (0.5 mg $kg^{-1}$ per day for 14 days) compared with saline-treated mice (Fig. 5f–h). Increased levels of circulating glucagon and insulin are a hallmark of the counter-regulatory response engaged during periods of fasting to maintain homeostatic blood glucose levels. The fact that nicotine-experienced rats showed counter-regulatory-like increases in glucagon and insulin, even though they were fed ad libitum and their body were similar to saline-experienced rats (Extended Data Fig. 9), suggests that nicotine alters metabolism to trigger hunger-like adaptions in the regulation of blood glucose. To investigate the role for TCF7L2 in this action, we examined the effects of chronic saline or nicotine (1 mg $kg^{-1}$ per day; 21 consecutive days) injections on blood glucose homeostasis in mutant and wild-type *Tcf7l2* rats (Fig. 5i). There were no differences in pre-treatment blood glucose or glucagon levels in mutant and wild-type *Tcf7l2* rats (Extended Data Fig. 9). After chronic saline treatment, fasting levels of blood glucose were again similar in wild-type and mutant *Tcf7l2* rats (Fig. 5j). However, chronic nicotine treatment increased fasting blood glucose (Fig. 5k) and glucagon (Extended Data Fig. 9) levels in wild-type compared with mutant *Tcf7l2* rats when assessed five and twelve weeks after the final injection (Fig. 5), which suggests that TCF7L2 regulates the emergence of diabetes-like abnormalities in blood glucose homeostasis in nicotine-experienced animals (Extended Data Fig. 10).

## Conclusions

Our findings reveal a crucial role for the diabetes-associated transcription factor TCF7L2 in regulating the function of nAChRs in the habenula and in controlling nicotine intake. Notably, we find that habenular neurons provide polysynaptic input to the pancreas and that nicotine acts on this habenula–pancreas circuit, in a TCF7L2-dependent manner and via the autonomic nervous system, to increase blood glucose levels. The reason why mHb stimulation should evoke such robust increases in blood glucose is unclear, but recent findings suggest that the mHb has a key role in coordinating adaptive responses to stressful or threatening stimuli[37–40]. Hence, it is likely that the mHb also regulates hyperglycaemic responses to stress, the function of which is to mobilize energy stores for 'fight-or-flight' behaviours[41]. By repeatedly hijacking this mHb-regulated stress response, chronic use of nicotine precipitates abnormalities in blood glucose homeostasis in a TCF7L2-dependent manner. It is notable that this pattern of effects is usually seen during periods of fasting, which suggests that nicotine withdrawal is associated with maladaptive alterations in metabolism that are analogous to a state of hunger. Loss of TCF7L2 function in the mHb increased nicotine consumption yet reduced hyperglycaemic responses to the drug and protected against abnormalities in blood glucose homeostasis. If these findings extend to human smokers, they would suggest a complex action in which deficits in TCF7L2 signalling increase the risk of tobacco dependence yet simultaneously protect against smoking-related diabetes. More broadly, our findings provide compelling evidence that diabetes and perhaps other smoking-related diseases, such as hypertension and cardiovascular disease, may originate in the brain and reflect nicotine-induced disruption of habenula-regulated interactions with the autonomic nervous system.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1653-x.

Received: 13 November 2018; Accepted: 12 September 2019;
Published online 16 October 2019.

1. Stolerman, I. P. & Jarvis, M. J. The scientific case that nicotine is addictive. *Psychopharmacology (Berl.)* **117**, 2–10, discussion 14–20 (1995).
2. Maskos, U. et al. Nicotine reinforcement and cognition restored by targeted expression of nicotinic receptors. *Nature* **436**, 103–107 (2005).
3. Fowler, C. D., Lu, Q., Johnson, P. M., Marks, M. J. & Kenny, P. J. Habenular α5 nicotinic receptor subunit signalling controls nicotine intake. *Nature* **471**, 597–601 (2011).
4. Fowler, C. D. & Kenny, P. J. Nicotine aversion: neurobiological mechanisms and relevance to tobacco dependence vulnerability. *Neuropharmacology* **76 Pt B**, 533–544 (2014).
5. Tuesta, L. M. et al. GLP-1 acts on habenular avoidance circuits to control nicotine intake. *Nat. Neurosci.* **20**, 708–716 (2017).
6. Liu, Z. & Habener, J. F. Glucagon-like peptide-1 activation of TCF7L2-dependent Wnt signaling enhances pancreatic beta cell proliferation. *J. Biol. Chem.* **283**, 8723–8735 (2008).
7. Grant, S. F. et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**, 320–323 (2006).
8. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
9. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
10. Haggard, H. W. & Greenberg, L. A. The effects of cigarette smoking upon the blood sugar. *Science* **79**, 165–166 (1934).
11. Sandberg, H., Roman, L., Zavodnick, J. & Kupers, N. The effect of smoking on serum somatotropin, immunoreactive insulin and blood glucose levels of young adult males. *J. Pharmacol. Exp. Ther.* **184**, 787–791 (1973).
12. GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet* **389**, 1885–1906 (2017).
13. Willi, C., Bodenmann, P., Ghali, W. A., Faris, P. D. & Cornuz, J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. *J. Am. Med. Assoc.* **298**, 2654–2664 (2007).
14. Ren, J. et al. Habenula "cholinergic" neurons co-release glutamate and acetylcholine and activate postsynaptic neurons via distinct transmission modes. *Neuron* **69**, 445–452 (2011).
15. Görlich, A. et al. Reexposure to nicotine during withdrawal increases the pacemaking activity of cholinergic habenular neurons. *Proc. Natl Acad. Sci. USA* **110**, 17077–17082 (2013).
16. Dougherty, J. D., Schmidt, E. F., Nakajima, M. & Heintz, N. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* **38**, 4218–4230 (2010).
17. Ables, J. L. et al. Retrograde inhibition by a specific subset of interpeduncular α5 nicotinic neurons regulates nicotine preference. *Proc. Natl Acad. Sci. USA* **114**, 13012–13017 (2017).
18. Boj, S. F. et al. Diabetes risk gene and Wnt effector *Tcf7l2*/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* **151**, 1595–1607 (2012).
19. Korinek, V. et al. Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* **19**, 379–383 (1998).
20. Geurts, A. M. et al. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* **325**, 433 (2009).
21. Graham, T. A., Ferkey, D. M., Mao, F., Kimelman, D. & Xu, W. Tcf4 can specifically recognize β-catenin using alternative conformations. *Nat. Struct. Biol.* **8**, 1048–1052 (2001).
22. Molenaar, M. et al. XTcf-3 transcription factor mediates beta-catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
23. Ip, W., Shao, W., Chiang, Y. T. & Jin, T. The Wnt signaling pathway effector TCF7L2 is upregulated by insulin and represses hepatic gluconeogenesis. *Am. J. Physiol. Endocrinol. Metab.* **303**, E1166–E1176 (2012).
24. Huang, S. M. et al. Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).
25. McGehee, D. S., Heath, M. J., Gelber, S., Devay, P. & Role, L. W. Nicotine enhancement of fast excitatory synaptic transmission in CNS by presynaptic receptors. *Science* **269**, 1692–1696 (1995).
26. Zoli, M., Léna, C., Picciotto, M. R. & Changeux, J. P. Identification of four classes of brain nicotinic receptors using β2 mutant mice. *J. Neurosci.* **18**, 4461–4472 (1998).
27. Murray, K. D., Choudary, P. V. & Jones, E. G. Nucleus- and cell-specific gene expression in monkey thalamus. *Proc. Natl Acad. Sci. USA* **104**, 1989–1994 (2007).
28. Skoglund, G., Hussain, M. A. & Holz, G. G. Glucagon-like peptide 1 stimulates insulin gene promoter activity by protein kinase A-independent activation of the rat insulin I gene cAMP response element. *Diabetes* **49**, 1156–1164 (2000).
29. Giniatullin, R., Nistri, A. & Yakel, J. L. Desensitization of nicotinic ACh receptors: shaping cholinergic signaling. *Trends Neurosci.* **28**, 371–378 (2005).
30. Paradiso, K. & Brehm, P. Long-term desensitization of nicotinic acetylcholine receptors is regulated via protein kinase A-mediated phosphorylation. *J. Neurosci.* **18**, 9227–9237 (1998).
31. Huganir, R. L., Delcour, A. H., Greengard, P. & Hess, G. P. Phosphorylation of the nicotinic acetylcholine receptor regulates its rate of desensitization. *Nature* **321**, 774–776 (1986).
32. Li, Y. F. et al. Antidepressant- and anxiolytic-like effects of the phosphodiesterase-4 inhibitor rolipram on behavior depend on cyclic AMP response element binding protein-mediated neurogenesis in the hippocampus. *Neuropsychopharmacology* **34**, 2404–2419 (2009).
33. Zhao, T. J. et al. Ghrelin secretion stimulated by $\beta_1$-adrenergic receptors in cultured ghrelinoma cells and in fasted mice. *Proc. Natl Acad. Sci. USA* **107**, 15868–15873 (2010).
34. Neil-Dwyer, G., Bartlett, J., McAinsh, J. & Cruickshank, J. M. β-adrenoceptor blockers and the blood–brian barrier. *Br. J. Clin. Pharmacol.* **11**, 549–553 (1981).
35. O'Dell, L. E. & Nazarian, A. Enhanced vulnerability to tobacco use in persons with diabetes: a behavioral and neurobiological framework. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **65**, 288–296 (2016).
36. Perkins, K. A., Epstein, L. H., Sexton, J. E. & Pastor, S. Effects of smoking cessation on consumption of alcohol and sweet, high-fat foods. *J. Subst. Abuse* **2**, 287–297 (1990).
37. Yamaguchi, T., Danjo, T., Pastan, I., Hikida, T. & Nakanishi, S. Distinct roles of segregated transmission of the septo-habenular pathway in anxiety and fear. *Neuron* **78**, 537–544 (2013).
38. Soria-Gómez, E. et al. Habenular CB1 receptors control the expression of aversive memories. *Neuron* **88**, 306–313 (2015).
39. Zhang, J. et al. Presynaptic excitation via GABA$_B$ receptors in habenula cholinergic neurons regulates fear memory expression. *Cell* **166**, 716–728 (2016).
40. Chou, M. Y. et al. Social conflict resolution regulated by two dorsal habenular subregions in zebrafish. *Science* **352**, 87–90 (2016).
41. Zhao, Z. et al. A central catecholaminergic circuit controls blood glucose levels during stress. *Neuron* **95**, 138–152.e5 (2017).

## METHODS

**Animals.** *Tcf7l2* mutant rats were generated by directing zinc-finger nucleases to a 169-base-pair (bp) region that spans the end portion of exon 5, which encodes the β-catenin binding domain, and the beginning portion of following intron on a Dahl/SS (derived from Sprague Dawley) background[20] (Extended Data Fig. 1). Accurate targeting of this region was confirmed by Sanger sequencing of genomic DNA from *Tcf7l2* mutant rats (Extended Data Fig. 1). This partial deletion of exon 5 is predicted to produce 2 truncated proteins, one that is 427 amino acids long and in which most of the β-catenin binding domain has been deleted, and the other a shorter protein (207 amino acids) that contains only the N-terminal region of the β-catenin binding domain. To detect deletion near the targeted binding site of zinc-finger nuclease, primers flanking exon 5 were used to amplify the intervening region by PCR from genomic DNA. The primer sequences were Tcf7l2-seq-F, CGCACAATGCTTATTCCTTAGC and Tcf7l2-seq-R, GGACGCCCACAAGTCTAGC. Wild-type and mutant PCR products were purified and sequenced by Sanger sequencing using the same primers. A 169-bp region comprising 49 bp of exon 5 and 120 bp of the following intron was confirmed to be deleted. Open reading frames of the predicted mutant *Tcf7l2* mRNA were generated using Vector NTI software and based on the NCBI Tcf7l2 transcript NM_001191052.1. Wild-type *Tcf7l2* and mutant *Tcf7l2* rats were obtained by heterozygous mating schemes. Mutant *Tcf7l2* rats are available on request. Rats were housed two per cage in an AALAC-approved vivarium on a 12-h reverse light–dark cycle. All experiments were carried out according to approved protocols from The Scripps Research Institute and Icahn School of Medicine at Mount Sinai Institutional Animal Care and Use Committees.

*Rosa26*$^{LSL\text{-}spCas9\text{-}eGFP}$ (026175), *Chat-cre* (006410), *Chat*$^{DW167}$ TRAP (030250), BAT-GAL (00531) and *Rosa-tdTom* (007914) mice were obtained from Jackson Laboratories and were bred in our animal facilities. All breeding was conducted by mating heterozygous pairs. Mice were housed in cages of 1–3 and were at least 6 weeks of age at the beginning of experiments.

**Drugs.** For self-administration experiments in mice and rats, nicotine bitartrate dihydrate (6019-06-3, MP Biomedicals) was dissolved in 0.9% sterile saline. All doses of nicotine refer to the free-base form. Atenolol (29122-68-7, Sigma-Aldrich,) was dissolved in 2 mM HCl in 0.9% sterile saline. The GLP-1 receptor agonist EX-4 (Tocris) was dissolved in 0.9% saline solution for in vivo use and 0.1 M PBS for ex vivo use. CNO (Enzo Life Sciences), insulin (11070-73-8, Sigma-Aldrich) and glucagon (16941-32-5, Sigma-Aldrich) were diluted in 0.9% saline for intraperitoneal injection. D-(+)-Glucose (50-99-7, Sigma-Aldrich) was dissolved in distilled $H_2O$ for oral gavage. cAMPS-RP, 8-Br-cAMP, and Rolipram (Tocris) were all dissolved in 0.9% saline. The pH of all solutions was adjusted to approximately 7.4.

**Cell culture.** Cell lines were maintained at 37 °C in a 5% $CO_2$ atmosphere. For experiments, cells were plated at $5 \times 10^4$ cells per well in tissue culture-treated 24-well plates (Corning) and transfected using the Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific) per the manufacturer's instructions unless otherwise noted. HEK293T cells expressing α4β2α5 nAChRs (gift from J. Lindstrom) were cultured in DMEM medium containing 25 mM HEPES, 2 mM glutamine, 1 mM pyruvate, 10% FBS (Gibco) and 1% penicillin/streptomycin (pen/strep). Rat INS-1 cells (a pancreatic β-cell line; gift from A. Stewart) were cultured in RPMI 1640 medium containing 25 mM HEPES, 2 mM glutamine, 1 mM pyruvate, 10% FBS (Gibco) and 1% pen/strep. Rat PC12 cells (a neuroblastic pheochromocytoma cells of the adrenal medulla) were grown in F-12k medium (Gibco) with 10% FBS, 5% horse serum and 1% pen/strep, and differentiated using nerve growth factor (NGF) stimulation (50 ng $ml^{-1}$). Mouse Neuro2A cells were grown in DMEM medium (Cellgro, Corning) containing 10% FBS and 1% pen/strep. For luciferase reporter assays, samples were processed using the Dual-Glo Luciferase Assay System (Promega) per the manufacturer's instructions. PC12 and Neuro2A cells lines were purchased from ATCC. Cells were not authenticated or tested for mycoplasma contamination.

**DNA vectors.** For luciferase experiments in INS-1 and PC12 cell lines, we obtained p-Lenti-7xTcf-FFluc-SV40-mCherry from Addgene (7TFC; Addgene plasmid 24307, gift from R. Nusse). The p-Lenti-7xTcf-FFluc-SV40-mCherry construct contained seven consensus binding sequences for human *TCF7L2* to control luciferase transcription. EVX1-CREB-Luciferase-GFP was a gift from M. Conkright. The EVX1-CREB-Luciferase-GFP contains cAMP-responsive elements to control luciferase transcription. A dominant-negative *Tcf7l2* construct (EdTc) was obtained from Addgene (Addgene plasmid 24310, gift from R. Nusse). The pGF-CREB-mCMV-EF1α-Puro CREB reporter and the control vector in a lentivirus backbone were purchased from System Biosciences (TR202va-p). The vector was packaged into active lentivirus particles (>$10^8$ infectious units, IFUs) for in vivo experiments.

**Viral vectors.** All viruses were distributed into 10-μl aliquots, kept at −80 °C, and thawed immediately before injection. For knockdown of *Glp1r* transcripts in rat brain, an shRNA construct that knocks down rat *Glp1r* by >80% in cultured cells and efficiently reduces *Glp1r* transcripts in rat brain (AAV1-shGlp1r-GFP; serotype 1) was used (gift from M. Hayes). The sequence of the shRNA was 5′-GATCGGGTTGCTGGTGGAAGGCGTGTATCTGTACTCAAGA GGTACAGATACACGCCTTCCACCAGCAACCTTTTTT-3′. Knockdown of *Glp1r* in the mHb was confirmed by qPCR using the rat *Glp1r* Taqman assay, Rn00562406_m1 (ThermoFisher Scientific). After injection, rats were allowed to recover for at least two weeks before experimentation.

**RNA extraction.** For RNA extraction and analysis, habenula tissues were homogenized in 500 μl TRIzol Reagent (ThermoFisher Scientific) according to the manufacturer's instructions. Samples were sonicated to complete homogenization. Chloroform (100 μl) was added and samples were vigorously vortexed for 15 s. After a 3-min incubation at room temperature, the samples were centrifuged at 12,000*g* at 4 °C for 15 min. The aqueous phase of the sample was removed by angling the tube at 45° and pipetting the solution to a fresh tube. One hundred per cent isopropanol (250 μl) was added to the aqueous phase and incubated for 10 min at room temperature. Samples were then centrifuged at 12,000*g* for 10 min at 4 °C. Supernatant was removed and discarded, leaving behind the RNA pellet. The pellet was washed with 500 μl 75% ethanol twice. The tube was allowed to dry for 5 min at room temperature. The pellet was resuspended in 20 μl nuclease-free water. Residual genomic DNA was removed using the DNA-free Kit (ThermoFisher Scientific) per the manufacturer's instructions. The concentration was measured using a NanoDrop machine (ThermoFisher Scientific). Samples were stored at −80 °C until processing.

**In vitro calcium measurement (FLIPR assays).** Cell lines were maintained at 37 °C in a 5% $CO_2$ atmosphere. For fluorometric imaging plate reader (FLIPR) assay, HEK293T cells were cultured in DMEM medium containing 25 mM HEPES, 2 mM glutamine, 1 mM pyruvate, 10% FBS (Gibco) and 1% pen/strep. Forty-eight and seventy-two hours after transfection with control siRNAs or siRNAs to knockdown targeted transcripts, cells were incubated with an equal volume of calcium-4 loading buffer (Molecular Devices) containing 2.5 mM probenecid at 37 °C for 30 min, followed by addition of vehicle or nicotine (dose range: 20 nM–320 μM) for another 30 min. The plates were then placed into a fluorometric imaging plate reader (Molecular Devices) to monitor fluorescence ($\lambda_{\text{excitation}} = 488$ nm, $\lambda_{\text{emission}} = 540$ nm).

**RT–PCR.** Samples were reverse transcribed into complementary DNA with the TaqMan High Capacity cDNA Reverse Transcription kit (ThermoFisher Scientific). Thereafter, they were processed with either the TaqMan Universal PCR kit with the rat *Glp1r* or *Tcf7l2* gene expression assay (ThermoFisher Scientific) or custom-made primers compatible with the Sybr Green Kit (ThermoFisher Scientific). Controls consisted of either *Actb* or *Gapdh*. Samples were quantified by qPCR (7900 Real-Time PCR system; ThermoFisher Scientific). All data were normalized relative to the mean housekeeping messenger RNA expression levels as an internal control. Comparison between groups was made using the method of $2^{-\Delta\Delta Ct}$.

**Brain perfusion and fixation.** Mice and rats were anaesthetized with an isoflurane (1–3%)/oxygen vapour mixture and perfused through the ascending aorta with 0.1 M PBS, followed by 4% paraformaldehyde (PFA) in 0.1 M PBS (pH 7.4). Brains were collected, post-fixed overnight in 4% PFA in 0.1 M PBS and then cryoprotected in 30% sucrose in 0.1 M PBS (pH 7.4) for 72 h at 4 °C. The cryoprotected brains were embedded in Tissue-Tek OCT compound (Finetek). Coronal sections (30–40 μm) were cut on a cryostat (Leica Biosystems) and collected directly onto slides and allowed to dry overnight at room temperature. Slides were stored at −20 °C until processing.

**Immunohistochemistry.** Sections (40 μm) containing the mHb from ChAT-tdTom mice were washed 3 times for 10 min in 0.1 M PBS (pH 7.4). Subsequently, sections were incubated in 10% normal donkey serum (NDS) with 0.5% Triton X-100 in PBS for 30 min at room temperature. The sections were then incubated with the TCF7L2 primary antibody (17-10109, Millipore Sigma) in 10% NDS, 0.5% Triton X-100 in PBS overnight at 4 °C. Slides were then washed 4 times for 10 min with 0.1 M PBS. Immunoreactivity was probed using Alexa-488/568-conjugated donkey anti-mouse/rabbit secondary antibodies (1:1,000; Molecular Probes) for 1 h at room temperature. The secondary antibodies were diluted in PBS-T containing 2% NDS. The slides were mounted with Fluoro-Gel containing DAPI (Electron Microscopy Sciences). The images were collected using a Zeiss AxioImager Z2 microscope system.

**Ex vivo cAMP stimulation and measurement of cAMP levels using ELISA.** Fresh habenula and IPn samples were micro-dissected from wild-type and mutant *Tcf7l2* rats. The tissue samples from two animals were pooled together for one experimental replicate. A total of three experimental replicates per condition were used. Tissue was lightly homogenized with a motorized tissue grinder in 40 μl PBS. To stimulate cAMP production in tissue homogenates, 5 μl of vehicle (PBS) or 1 mg $ml^{-1}$ EX-4 (final concentration 100 μM) was spiked into the tube. Samples were briefly mixed and incubated for 30 min at 30 °C. To stop the reaction, 5 μl of 1 M HCl was added to lyse the tissue. Samples were stored at −80 °C until use. When thawed, samples were centrifuged at 13,000*g* for 10 min at 4 °C. The concentration of cAMP in the supernatant obtained from the rat habenula and the IPn extracts was measured using the cAMP Direct Immunoassay Kit (Biovision) per the manufacturer's instructions.

**Ex vivo assessment of cAMP signalling using luciferase imaging.** On the day of each experiment, wild-type and mutant *Tcf7l2* rats that received intra-mHb injections with a lentivirus to express the pGF-CREB-mCMV-dscGFP-P2A-luciferase reporter were injected with luciferin (150 mg kg$^{-1}$, intraperitoneally) dissolved in sterile 0.9% saline (25 mg ml$^{-1}$). After 20 min, rats were lightly anaesthetized with isoflurane, decapitated and brains rapidly dissected on ice. Coronal slices (about 1.5 mm; 2–3 in total) were collected from each animal to include the full extent of the mHb. Slices were transferred to a 6-well plate and placed in a solution of 2.0–2.5 ml of oxygenated aCSF. All imaging was performed using the IVIS Spectrum imaging module from Caliper Life Sciences, with the manufacturer-provided Living Image software set to the following parameters: 4 min exposure length, medium binning, luminescent F/Stop set at 1, excitation filter blocked, and the emission filter open. Six-well plates with slices were placed in the IVIS and situated in the middle of the imaging grid with the stage temperature set to 37 °C. Imaging experiments began following an addition of luciferin (at a concentration of 0.3 mg ml$^{-1}$) to characterize 'baseline' luminance. To stimulate CREB activity, EX-4 (final concentration 250 nM) was then added to the bath and reads were collected every 5 min over the next 35 min. A region of interest (ROI) was manually drawn around the mHb in each sample, with the size of the ROI kept constant across all slices analysed. A total of 6 ROIs from wild-type rats (from 3 rats) and 7 from mutant rats (from 4 rats) were used for statistical comparison. Luminance was assessed over 35 min after bath application of EX-4 (250 nM).

**BAT-GAL transgenic mice and LacZ staining.** Adult (10–12-week-old) BAT-GAL transgenic were purchased from Jackson laboratories. The transgene expresses the *lacZ* gene under the control of a regulatory sequence consisting of seven consensus TCF7L2-binding motifs upstream of the *Xenopus siamois* gene minimal promoter. Transgenic mice display β-galactosidase activity in the presence of active transcription factor binding. Mice were perfused with 2% PFA in PBS and brains post-fixed in 2% PFA in PBS containing 2 mM $MgCl_2$ and 2 mM EGTA for 1 h at 4 °C. Brains were washed in wash buffer (0.1 M PBS, 2 mM MgCl2) 3 times for 15 min. Then, brains were cryoprotected in 15% followed by 30% sucrose in wash buffer. The cryoprotected brains were embedded in Tissue-Tek OCT compound (Finetek). Coronal sections (20 μm) containing the habenula were collected directly onto slides and dried overnight at room temperature. Slides were washed in wash buffer 10 min on ice. Slices were then washed in LacZ buffer (0.1M PBS, 2 mM $MgCl_2$, 0.01% sodium deoxycholate, 0.02% NP-40). Slides were placed in a humidified chamber and 200 μl LacZ stain (0.1 M PBS, 2 mM MgCl2, 0.01% sodium deoxycholate, 0.02% NP-40, 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide) was applied to each slide. Slides were placed at 37 °C for 2–3 h. After incubation, slides were washed in wash buffer twice for 5 min followed by a finally wash in distilled water for 5 min. Tissue was counterstained with Nuclear Fast Red (Vector Laboratories). Finally, slides were dehydrated through a methanol series (1 × 5 min each 50%, 70% and 100%), cleared in xylene twice for 5 min, and mounted with Permount (Fisher Scientific). Images of β-galactosidase staining were obtained by brightfield microscopy.

**$^{86}Rb^+$ efflux.** $^{86}$RbCl (average initial specific activity 15 Ci mg$^{-1}$) and Optiphase Supermix scintillation cocktail were purchased (Perkin-Elmer NEN). To obtain crude synaptosomal preparations of habenula and IPn, fresh tissue was microdissected and prepared as previously described[3]. Samples were loaded with $^{86}Rb^+$ and acetylcholine-stimulated $^{86}Rb^+$ efflux was measured, with each sample stimulated only once (see ref.[3]). $^{86}Rb^+$ efflux was expressed as the increase in signal above basal efflux. A nonlinear least-squares curve fit to a first-order equation ($Ct = C0 \times e^{-kt}$), in which $Ct$ is the basal efflux counts at time $t$, $C0$ is the estimated efflux counts at $t = 0$ s, and $k$ is the first-order decay constant) was used to estimate basal efflux for each sample. Counts in fractions preceding and following the peak were used for curve fitting. Acetylcholine-stimulated efflux was calculated by summing the counts in the fractions exceeding basal efflux during acetylcholine exposure and dividing by the corresponding basal efflux counts. This value represents total peak relative to baseline.

**Brain slice preparation for physiology recordings.** Three- to six-month-old male and female wild-type and mutant *Tcf7l2* rats were used for all electrophysiology experiments. Rats were anaesthetized with isoflurane followed by transcardial perfusion with oxygenated (95% $O_2$, 5% $CO_2$) *N*-methyl-D-glucamine (NMDG) HPEPS solution (in mM: 92 NMDG, 2.5 KCl, 1.2 $NaH_2PO_4$, 30 $NaHCO_3$, 20 HEPES, 25 glucose, 5 $Na^+$ ascorbate, 2 thiourea, 3 $Na^+$ pyruvate, 10 $MgSO_4{\cdot}7H_2O$, 0.5 $CaCl_2{\cdot}2H_2O$, with pH adjusted to 7.3–7.4, 300–310 mOsm). The brain was quickly removed into ice-cold NMDG HEPES solution for 1 min. For nAChR desensitization studies, 300-μm-thick coronal slices containing MHb were cut with a vibratome (Leica VT1200S). For studies involving measurement of sEPSCs in the IPn, a custom brain block was used to cut the brain at an approximately 55° angle before collecting 300–350-μm-thick angled coronal slices containing intact mHb-Fr-IPn circuitry cut by a vibratome. Slices were then incubated at 32 °C for 25–35 min, then kept at room temperature for at least 1 h, in 95% $O_2$/5% $CO_2$-equilibrated HEPES-holding-solution, containing (in mM): 92 NaCl, 2.5 KCl, 1.2 $NaH_2PO_4$, 30 $NaHCO_3$, 20 HEPES, 25 glucose, 5 $Na^+$ ascorbate, 2 thiourea, 3 $Na^+$ pyruvate, 2 $MgSO_4{\cdot}7H_2O$, and 2 $CaCl_2{\cdot}2H_2O$.

**Voltage-clamp electrophysiology.** Recordings were made under an upright microscope (Scientifica SliceScope Pro 2000) equipped with infrared differential interference contrast optics for visualization. Slices were transferred to a recording chamber superfused with standard recording ACSF containing (in mM): 124 NaCl, 2.5 KCl, 1.2 $NaH_2PO_4$, 24 $NaHCO_3$, 5 HEPES, 12.5 glucose, 2 $MgSO_4{\cdot}7H_2O$ and 2 $CaCl_2{\cdot}2H_2O$, adjusted to pH 7.3–7.4, 295–305 mOsm. Recordings were performed at 32 °C. Patch pipettes were made from borosilicate glass capillary tubing (1B150F-4; World Precision Instruments) using a micropipette puller (PC-10; Narishige).

For nAChR desensitization studies, the internal recording pipette solution was potassium-based and contained the following (in mM): 130 $K^+$ gluconate, 4 KCl, 0.3 EGTA, 10 HEPES, 4 MgATP, 0.3 $Na_2GTP$, 10 phosphocreatine; pH adjusted to 7.3 with KOH while the external solution was ACSF plus 0.5 μM tetrodotoxin, 100 μM picrotoxin, 5 μM NBQX and 50 μM 2-amino-5-phosphonovaleric acid (AP5). nAChR currents were recorded from the soma of mHb neurons using a Multiclamp 700B amplifier (Molecular Devices), filtered at 3 kHz, amplified 5 times, and then digitized at 10 kHz with a Digidata 1550 analogue-to-digital converter (Molecular Devices). Voltage was held at −60 mV ($V_{hold} = -60$ mV). A baseline nAChR current was recorded for 3 min. For drug application, a 30 μM nicotine-filled glass pipette, identical to a typical recording pipette, was connected to a micropressure ejection system (PICOSPRTIZERIII, Parker). Ejection pipettes were moved to within 20–40 μm of the recorded cell using a manipulator, drug was applied for 0.1 Hz. For the induction of nAChR desensitization, the frequency of drug application was changed to 1 Hz for 60 s. After the induction, the frequency was returned to 0.1 Hz and nAChR were continuously measured for another 10 min.

For sEPSC measurements in the IPn, the internal recording pipette solution was potassium-based contained the following (in mM): 130 $K^+$ gluconate, 4 KCl, 0.3 EGTA, 10 HEPES, 4 MgATP, 0.3 $Na_2GTP$, 10 phosphocreatine; pH adjusted to 7.3 with KOH and the external solution was ACSF plus+ 100 μM picrotoxin. sEPSCs were recorded from the IPn neurons using a Multiclamp 700B amplifier with a DigiData 1500 interface and Clampex 10.3 software (Molecular Devices) sampled at 5 kHz and low-pass-filtered at 1 kHz. Voltage was held at −70 mV ($V_{hold} = -70$ mV). A baseline sEPSC was recorded for 5 min. For drug application, 50 μM nicotine was locally applied to ventral MHb by gravity using nicotine-filled glass pipette (about 200 μm) and was sucked by a local suction glass pipette (about 200 μm) that was positioned to outside of mHb. The local suction pipette prevented the diffusion of nicotine to the recording site in the IPn. Local perfusion pipettes were moved to within 100–200 μm of the brain slice surface using a manipulator, drug was applied continuously for 5 min with a local suction pipette and sEPSCs were continuously measured for another 5 min.

**Functional magnetic resonance imaging.** Rats were prepared with a tail vein catheter immediately before being placed in the scanner. Rats were first anaesthetized using isoflurane anaesthesia (3% induction and 1.5% maintenance), then a bolus of 0.05 mg kg$^{-1}$ medetomindine was administered subcutaneously. Isoflurane was discontinued 5 min after the bolus administration. Medetomidine was then infused (0.1 mg kg$^{-1}$ per h) via the tail vein catheter to maintain sedation. All rats were imaged on a heated bed and respiration was monitored continuously until the end of the scan. After anaesthesia was established, functional magnetic resonance imaging acquisition began. All imaging was performed using a Bruker Biospec 70/30 7 Tesla scanner with a B-GA12S gradient insert (gradient strength 440 mT m$^{-1}$ and slew rate 3,444 T m$^{-1}$ s$^{-1}$). A Bruker 4 channel rat brain phased array was used for all data acquisition in conjunction with a Bruker volume transmit 86-mm coil. After a three-plane localizer, three short anatomical T2 scans were acquired for the purpose of anatomical localization and co-registration. Functional scans were acquired with a GE-EPI protocol with the following parameters: TR = 3,000 ms, TE = 15 ms, flip angle = 70°, field of view = 25.6 mm × 25.6 mm, matrix size = 80 × 80, in-plane resolution = 320 μm × 320 μm, number of slices = 38, slice thickness = 0.7 mm, slice gap = 0 mm, number of volumes = 800, and dummy scans = 4. The total scanning time for the functional magnetic resonance imaging experiment was 40 min. After 20 min of acquisition, rats were injected with CNO (1 mg kg$^{-1}$) intravenously and data acquisition continued for another 20 min for a 40 min scanning session in total.

**RNA-seq, TRAP and differential analysis.** RNA-seq data generated from Illumina HiSeq 2500 were processed following an open source pipeline[42]. A total of $n = 9$ rats of each genotype (wild-type and mutant *Tcf7l2*) were used. RNA libraries for each brain region were generated from pooled RNA from three animals per group. In brief, the paired-end sequencing reads were aligned to the human genome (version hg19) and rat genome (version rn6), using the Spliced Transcripts Alignment to a (STAR)[43]. Next, featureCount[44] was used to assign aligned reads to genes. Counts per million were used as the expression quantification method. The counts per million matrix was $\log_2$-transformed and the $Z$-score was scaled to centre the expression values of each gene to 0 with a standard deviation of 1 before performing principal component analysis and hierarchical clustering. The characteristic direction[45] was used to identify differentially expressed genes between the wild-type and mutant *Tcf7l2* samples. Enrichment analyses for differentially expressed genes were performed using Enrichr[46,47]. RNA-seq data generated from TRAP were processed and analysed as previously described[16,17].

***Tcf7l2* sgRNA synthesis and validation.** Five sgRNA sequences that targeted all known 17 murine transcript variants of *Tcf7l2* were designed and synthesized using *Streptococcus pyogenes* Cas9 protospacer adjacent motif (PAM) sites. Off-target cleavage was bioinformatically assessed using http://crispor.tefor.net. sgRNAs were synthesized via PCR with loci-specific primers followed by T7 in vitro transcription, purification and quantitation using the GeneArt Precision sgRNA Synthesis Kit from ThermoFisher (A29377). The N2a neuroblastoma cells at 60% confluency were transfected, via Lipofectamine MessengerMAX Transfection Reagent, with Cas9 (0.5 μg) and sgRNA (125 ng) in triplicate. Cells were incubated for 48 h, lysed and PCR was carried out using target-specific primers. The PCR strands were re-annealed and mismatches digested. Mismatches were quantified using a T7 endonuclease-based mismatch assay (ThermoFisher A24372). The sgRNA with the greatest cleavage efficiency for *Tcf7l2*, and the fewest predicted off-targets, was selected for in vivo testing; sequence: GTGTACCCAATCACGACAGGAGG. The predicted off targets of this sgRNA were intergenic only. The selected *Tcf7l2* sgRNA was cloned into an AAV plasmid (AAV:ITR-U6-sgRNA(backbone)-pCBh-Cre-WPRE-hGHpA-ITR) (Addgene plasmid 60229). The sgRNA backbone was replaced with the sgRNA sequence for *Tcf7l2* or eGFP to serve as a control vector; sequence: GAGCTGGACGGCGACGTAAACGG. The Cre recombinase cassette in the plasmid was replaced with dTomato by Vector Biolabs. The plasmids were packaged into infectious particles with a titre greater than $1.0 \times 10^{13}$ and aliquoted into 5 μl volumes and stored at −80 °C until use. The AAV carrying the sgRNA for *Tcf7l2* was validated by viral transduction with AAV-CMV-spCas9 (Vector Biolabs) in N2a cells. Fluorophore abundance, mRNA knockdown and genomic cleavage efficiency were assessed in vitro before infection. To confirm in vivo cleavage of *Tcf7l2*, *Rosa26*$^{LSL\text{-}spCas9\text{-}eGFP}$ mice were killed by cervical dislocation, their brains were removed and frozen immediately in methylbutane on dry ice. The mHb was dissected using a 1.0-mm punch on the cryostat. Genomic DNA from the mHb was extracted using Purelink genomic DNA kit (k1820-01) from ThermoFisher Scientific. Primers were designed to surround the CRISPR cut site. forward: AGCTTACTGTACGGCGAGAAC and reverse: TGTCTAGGTGAGTCGCTGTG. DNA amplicons were generated by PCR. The PCR product was purified using Qiagen PCR purification kit (28104). DNA amplicons were sequenced using EZ amplicon sequencing (Genewiz) and the percentage of indels calculated.

**Cannula implantation and intracranial microinjection.** For CRISPR-mediated cleavage of *Tcf7l2*, AAV2-hSYN1-iCre-WPRE and AAV 1-U6-sgRNA (Tcf7l2)-pCBh-dTomato-WPRE-hGHpA-ITR or AAV 1-U6-sgRNA (eGFP)-pCBh-dTomato-WPRE-hGHpA-ITR were locally infused into the mHb of male *Rosa26*$^{LSL\text{-}spCas9\text{-}eGFP}$ mice. The coordinates from bregma were: anteroposterior (AP) −1.55, mediolateral (ML) 1.65, from dura dorsoventral (DV) −2.9 to −3.1, at an angle of 32°. The incubation period was three months before behaviour commenced. mRNA levels (Thermo Fisher Taqman probe: Mm01258049_m1) and genomic cleavage (Thermo Fisher A24372) were quantified to ensure effective genomic editing of *Tcf7l2*. For intra-mHb siRNA injections in rats, three pooled siRNAs (ON-TARGETplus siRNA, GE Healthcare Bio-Sciences, LQ-107966) were purchased and diluted to 1 mg ml$^{-1}$ with nuclease-free water. To deliver siRNA to the mHb, we used the jetSI 10 mM reagent (Polyplus-transfection) and followed the manufacturer's directions with some minor modifications. To prepare siRNA duplexes for in vivo delivery, 2.5 μl 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE; Sigma-Aldrich) was added to 10 μl of 10 mM jetSI. To this, 7.5 μl of 100% molecular-grade ethanol (Sigma-Aldrich) was added to create reagent B. Then, 3 μl of reagent B was added to 7.5 μl filter-sterilized 25% glucose and 27 μl nuclease-free water to create reagent C. Reagent C was vortexed vigorously and incubated at room temperature for 10 min. Separately, the siRNA tube was prepared. To this tube, 2.8 μl of siRNA (control or *Tcf7l2*), 1.25 μl 25% filter-sterilized glucose and 2.2 μl nuclease-free water were added for a total volume of 6.25 μl and vortexed gently. Next, 6.25 μl of reagent C was added to the siRNA tube preparation and immediately vortexed for 10 s. The mixture was incubated for 30 min at room temperature to allow siRNA duplex formation. After this incubation, the prepared solution was injected into animals within 30 min. One microlitre (approximately 0.22 μg μl$^{-1}$) of siRNA duplex was infused through bilateral cannulae–injector system (PlasticsOne) directed towards the mHb as described above. To knock down *Glp1r* transcripts in the mHb, rats were anaesthetized with an isoflurane (1–3%) and oxygen vapour mixture and positioned in a stereotaxic frame (Kopf Instruments). The incisor bar was set to the 'flat-skull' position. Rats were injected with AAV1-shGlp1r-GFP or AAV1-GFP virus particles (titre $= 5 \times 10^{12}$) according to the following stereotaxic coordinates: for bilateral mHb injections; flat skull, 10° angle towards midline; AP: 3.2 mm from bregma; ML: ±1.35 mm from midline; DV: −5.3 mm from skull surface. During microinjections, the injector needles extended into mHb and virus particles were administered in a volume of 0.3 μl and at a rate of 0.1 μl min$^{-1}$. The injector needle remained in place for 2 min after injection. Animals were allowed to recover for at least 72 h, during which time they were administered post-surgery antibiotic and analgesic. For intra-mHb drug infusions, guide cannulae (PlasticsOne) were implanted as follows: flat skull; 10° angle towards midline; AP: −3.2 mm from bregma; ML: ±1.4 mm from midline; DV: −3.3 mm from skull surface. Guide cannulae were encased in dental acrylic to attach to the skull surface and hold in place. Animals were allowed to recover for at least 72 h, during which time they were administered post-surgery antibiotic and analgesic. During microinjections, the injector system (PlasticsOne) was designed to extend 2 mm below the tip of the cannula for placement into the mHb. Here, EX-4 (or vehicle) was administered in a volume of 0.5 μl and at a rate of 0.333 μl min$^{-1}$. The injector needle remained in place for a minimum of 2 min after injection to allow for diffusion and to prevent backflow into the cannulae. EX-4 was dissolved in sterile, 0.1 M PBS. For intra-IPn drug infusions (DKK1 and XAV939), rats were anaesthetized with an isoflurane (1–3%) and oxygen vapour mixture and guide cannulae (PlasticsOne) were implanted as follows: flat skull; 10° angle towards midline; AP: −5.2 mm from bregma; ML: +1.5 mm from midline; DV: −7.2 mm from skull surface. Guide cannula were encased in dental acrylic to attach to the skull surface and hold in place. Animals were allowed to recover for at least 72 h, during which time they were administered post-surgery antibiotic and analgesic. During microinjections of DKK1 and XAV939, the injector needles were designed to extend 2 mm below the tip of the cannula for placement into the IPn and drug was administered in a volume of 0.5 μl and at a rate of 0.333 μl min$^{-1}$ 30 min before the start of the experiment (100 ng μl$^{-1}$ DKK1 and 10 μg μl$^{-1}$ XAV939). The injector needle remained in place for a minimum of 2 min after injection to allow for diffusion and to prevent backflow into the cannula. All drugs were all dissolved in 0.1 M PBS unless otherwise noted. For ex vivo CREB luciferase imaging, wild-type and mutant *Tcf7l2* male rats were anaesthetized with 1–3% isoflurane and positioned in a stereotaxic frame (Kopf Instruments). Each rat received a bilateral injection of the CREB-reporter (PGF1-CREB; 300 nl) virus targeting the medial habenula (coordinates from bregma: −3.6 mm AP, ± 1.2 ML, −4.5 mm DV from the surface of the brain. The injector was set at 10° from midline). Animals were permitted at least four weeks to recover and allow time for reporter expression before ex vivo imaging experiments.

**Open field.** Rats were placed in an 80 × 80 cm Plexiglas chamber, and distance travelled was measured by live automated tracking using an overhead CCD camera and Ethovision software (Noldus Information Technology). Sessions were 60 min long, and all testing was performed during the animals' active/dark phase under red light. Rats were first tested naive to the open field. On each successive day, 15 min before testing they were injected subcutaneously with saline (days 1, 2, 6, 7) or 0.4 mg kg$^{-1}$ nicotine (days 3, 4, 5). One cohort of animals was tested before and one cohort was tested after undergoing intravenous nicotine self-administration.

**Intravenous nicotine self-administration.** Mice and rats were mildly food restricted to 85–90% of their free-feeding body weight and trained to press a lever in an operant chamber (Med Associates) for food pellets (20 mg pellets mice; 45 mg food pellets rats; TestDiet) under a fixed-ratio 5, time out 20 s (FR5TO20s) schedule of reinforcement before catheter implantation. Once stable responding was achieved (>20 pellets per session in mice; >50 pellets per session in rats), subjects were catheterized as described above. The animals were allowed at least 48 h to recover from surgery, then permitted to respond to food reinforcement again under the FR5TO20s schedule. Once food responding criterion was re-established, subjects were permitted to acquire intravenous nicotine self-administration by autoshaping during 1 h daily sessions, 7 days per week. Nicotine was delivered through the tubing into the intravenous catheter by a Razel syringe pump (Med Associates). Each nicotine self-administration session was performed using 2 retractable levers (1 active and 1 inactive) that extend 1 cm into the chamber. Completion of the response criteria on the active lever resulted in the delivery of an intravenous nicotine infusion (0.03 ml infusion volume for mice; 0.1 ml for rats). Responses on the inactive lever were recorded but had no scheduled consequences. For dose–response studies, animals were presented with each dose of nicotine for at least three days; the mean intake over the last two sessions for each dose was calculated and used for statistical analysis. Nicotine doses were presented according to a within-subjects Latin square design. In between each dose, subjects were placed back on the training dose for at least two days or until their intake returned to baseline levels before being tested on the next dose in the Latin-square design.

**Blood sampling and collection.** For blood glucose sampling (self-administration, oral glucose and fasting experiments), the Contour Blood Glucose Monitoring System (Bayer HealthCare) including meter and testing strips was used. The site of blood collection (end of the tail) was sterilized with 70% ethanol. A small blood sample, approximately 0.5 μl, was obtained by nicking the tail vein with a 22-gauge needle. After blood sampling, the site was cleaned with 70% ethanol. For ELISAs, the end of the tail was sterilized with 70% ethanol and nicked with a razor blade and around 300 μl blood was collected in tubes containing EDTA (RAM Scientific) and

500 kallikrein inhibitor unit per ml aprotinin (Sigma-Aldrich). To obtain serum, blood samples were allowed to clot for 30 min on a rocker, spun at 1,000*g* at 4 °C for 10 min and aliquoted into separate tubes. Samples that displayed notable haemolysis were excluded from analysis. Samples were stored at −80 °C until processing.

**Retrograde tracing with pRV-GFP.** We used retrograde neuronal tracer pseudorabies virus expressing green fluorescent protein (pRV-GFP) to identify CNS regions innervating pancreas. Stock of pRV-152 encoding GFP210 was prepared to concentrations of $3.4 \times 10^8$ plaque-forming units per ml and stored at −80 °C until use. Injection of pRV-152 (500 nl) was performed in anesthetized mice over 30 s using a Hamilton syringe fitted with a 30G1/2 needle, with the needle was left in place for 1 min after injection to allow for diffusion.

**Fasting blood sugar test.** Rats were food restricted for 4 or 16 h, with water maintained ad libitum. To sample fasting blood sugar levels, a small blood sample was obtained as described above.

**Oral glucose tolerance test.** Before the oral glucose tolerance test, rats were food restricted for 16 h, with water ad libitum. A baseline blood sample was obtained at the end of the fasting period, as described above. A bolus of glucose (1 g $\mathrm{kg}^{-1}$; 40% glucose solution, in water) was administered to rats by oral gavage. Blood sugar was sampled at 30, 60 and 120 min after the glucose bolus.

**Measurements of circulating glucagon and insulin.** Blood collection was performed as described above. Levels of glucagon and insulin were measured from serum using the Mouse/Rat Total Active Insulin/Glucagon Multi-Spot Assay System with a Sector Imager 2400 (Meso Scale Discovery) per the manufacturer's instructions.

**Statistical analyses.** Animal sample size was justified by previously published data or preliminary experiments. Data distribution was assumed to be normal, but this was not formally tested. In all experiments animals with the same genotype were randomly allocated to experimental groups. The investigators were not blinded to allocation during experiments and outcome assessment. For self-administration experiments, animals that did not achieve stable levels of intake ($<20\%$ variation in intake across 3 consecutive days) or that took less than five nicotine infusions on average across sessions were excluded from experiments. Data collected from animals after their catheters broke, but not before, were excluded from statistical analyses. Male and female wild-type and mutant *Tcf7l2* rats were used in nicotine self-administration experiment. All data were analysed two-sided *t*-test or by one-, two- or three-way ANOVA as appropriate using GraphPad Prism software. Significant main or interaction effects were followed by Bonferroni post hoc tests as appropriate. The criterion for significance was set at $P < 0.05$. For all electrophysiological data, results are shown as the mean ± s.e.m.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The RNA-seq data generated in this study are available at the Gene Expression Omnibus (GEO) under the accession code GSE137118. Other data that support the findings of this study are available as Extended Data and Supplementary Information, including uncropped western blot images.

42. Wang, Z. & Ma'ayan, A. An open RNA-Seq data analysis pipeline tutorial with an example of reprocessing data from a recent Zika virus study *F100Res.* **5**, 1574 (2016).
43. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
44. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
45. Clark, N. R. et al. The characteristic direction: a geometrical approach to identify differentially expressed genes. *BMC Bioinformatics* **15**, 79 (2014).
46. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
47. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44** (W1), W90–W97 (2016).

**Acknowledgements** This work was supported by grants from the National Institute on Drug Abuse to P.J.K. (DA020686) and I.I-T. (DA035756). We thank S. Stanley for pRV-GFP virus, J. Lindstrom for the α4β2α5 nAChR cell line, M. Conkright for the EVX1-CREB-luciferase-GFP reporter, A. Stewart for INS-1 cells and M. Hayes for the AAV1-shGlp1r-GFP virus. The Lenti-7xTcf-FFluc-SV40-mCherry reporter (7TFC) and dominant-negative *Tcf7l2* construct (EdTc) were gifts from R. Nusse. Oxycodone and cocaine were supplied by the NIDA Drug Supply Program.

**Author contributions** A.D., M.P.H., M.V.M.DiB., S.P.B.C., W.M.H., P.B., R.M.O.C., M.W., C.F. and K.S.E. performed all behavioural experiments; M.I. performed electrophysiological recordings; S.P.B.C. designed and validated gRNAs; X.L. and Z.C. performed virus tracing; Q.L. and T.M.K. performed cell culture experiments; J.L.A. and I.I-T. generated and analysed the TRAP data; A.D. and Q.L. generated the RNA-seq data from wild-type and mutant rats; Z.W. and A.M. analysed RNA-seq data; H.C.O'N. performed the rubidium efflux experiments; A.M.G. generated the *Tcf7l2* mutant rats; P.J.K. designed the experiments; A.D. and P.J.K. analysed the data and wrote the manuscript.

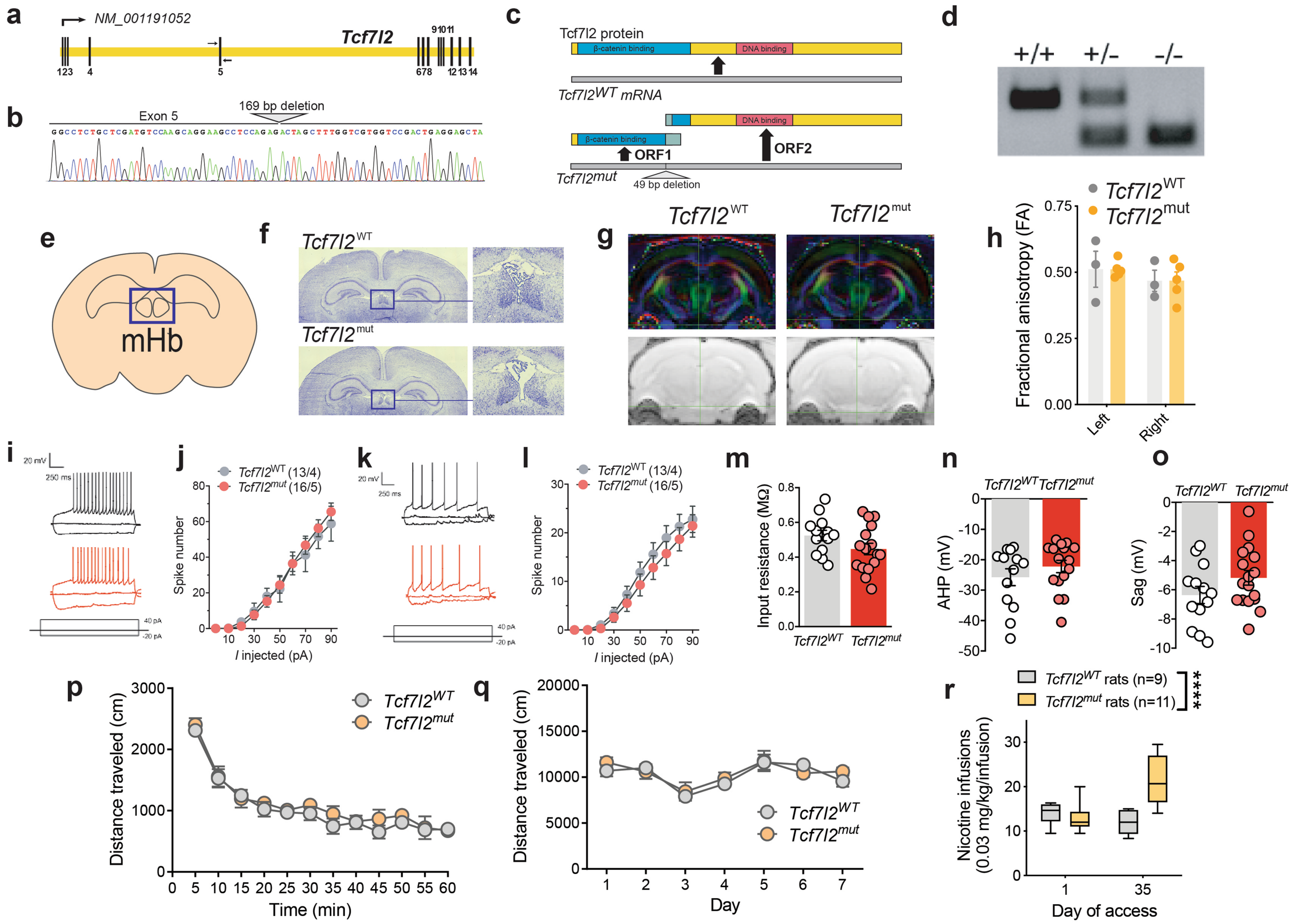
**Competing interests** P.J.K. is co-founder of Eolas Therapeutics Inc., which has a licensing agreement with AstraZeneca to develop small molecule treatments for drug dependence. P.J.K. has research support from Eli Lilly and Takeda USA.

### Additional information

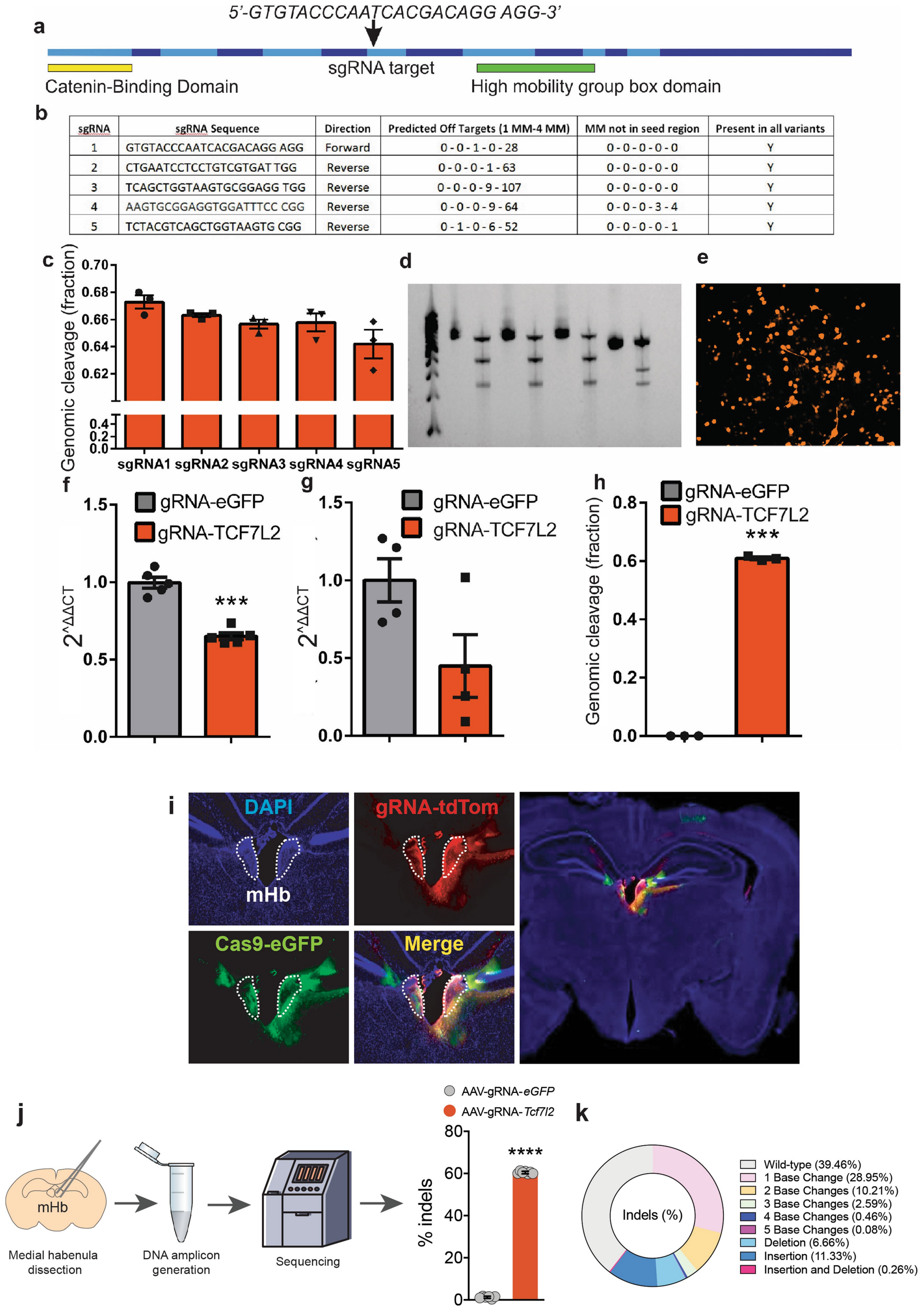
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1653-x.

**Correspondence and requests for materials** should be addressed to P.J.K.

**Peer review information** *Nature* thanks Peter Kalivas, Tamas Horvath and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



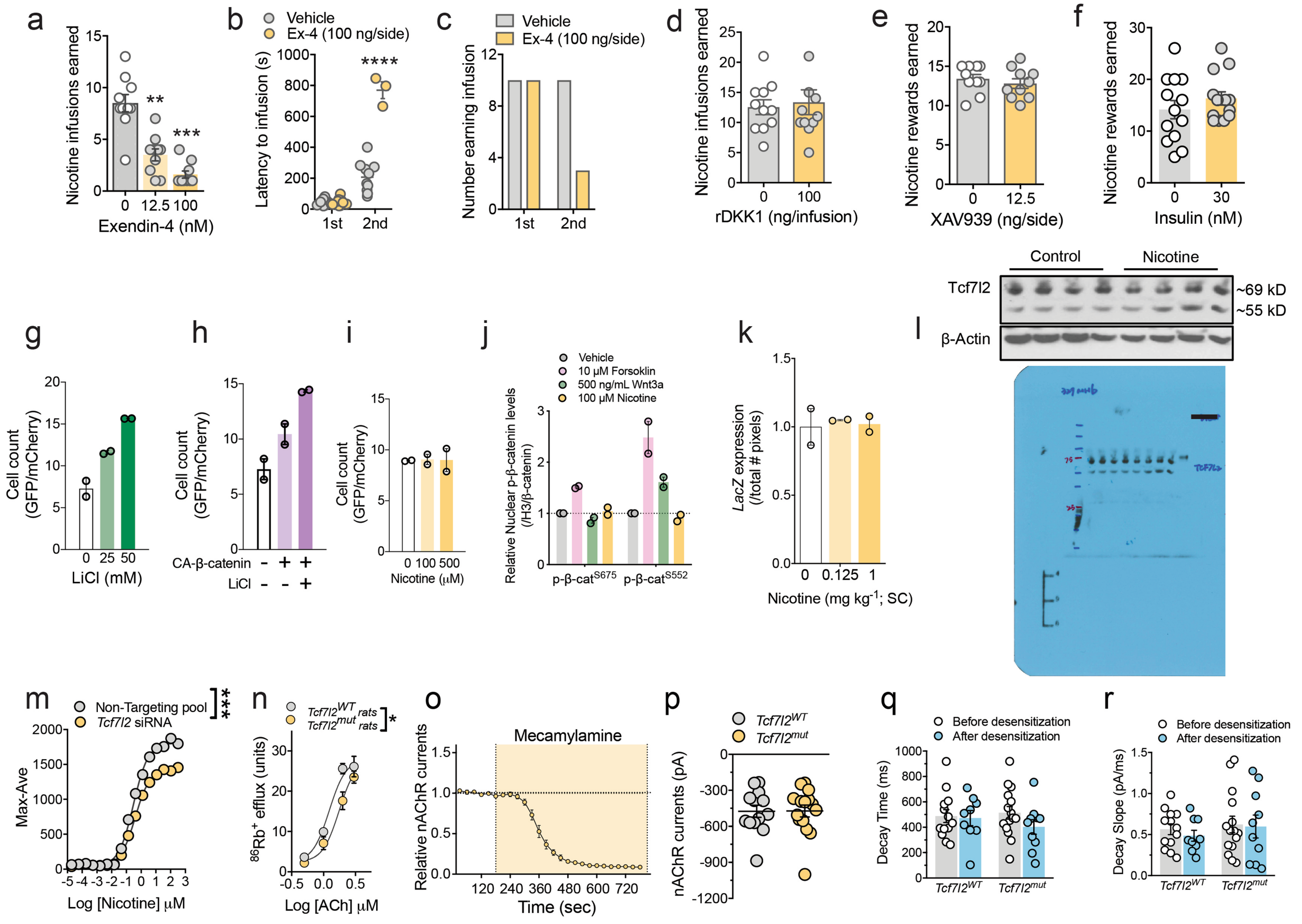
**Extended Data Fig. 1 | Generation of *Tcf7l2* mutant rats. a**, Schematic of the *Rattus norvegicus Tcf7l2* gene. Exons are spliced to generate *Tcf7l2* mRNA (NCBI reference sequence: NM_001191052.1). Primers for genotyping and Sanger sequencing are indicated by arrows flanking exon 5. **b**, Sequencing chromatograph of the *Tcf7l2* mutant allele. The site of the 169-bp deletion from exon 5 and the following intron is labelled. **c**, Illustration of TCF7L2 wild-type protein, containing an N-terminal β-catenin binding domain (blue) and C-terminal DNA binding domain (red). Predicted open reading frames and truncated proteins generated from the *Tcf7l2* mutant mRNA. Green regions on predicted truncated proteins denote ectopic amino acid sequences not found in wild-type TCF7L2 protein. **d**, Genotyping of wild-type and mutant *Tcf7l2* rats: wild-type animal (+/+) with single band at 304 bp; heterozygous animal (+/−) with bands at 304 and 144 bp; and mutant animal (−/−) with a single band at 144 bp. Image is representative of genotyping results obtained for wild-type and mutant *Tcf7l2* rats used each experiment. **e**, Graphical representation of mHb in coronal slice of rat brain. Image adapted from the Allen Brain Reference Atlas. **f**, Nissl staining showed similar mHb volumes in wild-type and mutant *Tc7l2* rats. Image is representative of results obtained in three biologically independent animals from each genotype. **g**, Diffusion tensor imaging tractography of the fasciculus retroflexus in wild-type ($n = 3$) and mutant ($n = 5$) *Tcf7l2* rats. **h**, Fractional anisotropy showed similar integrity (left and right sides) of the fasciculus retroflexus in wild-type ($n = 3$) and mutant ($n = 5$) *Tcf7l2* rats ('genotype': $F_{1,6} = 0.000003$; $P = 0.99$; 'brain side': $F_{1,6} = 2.562$, $P = 0.16$; 'genotype × brain side': $F_{1,6} = 0.0007$, $P = 0.98$). **i**, The frequency at different steps of positive current used to calculate the slope of the input–output curve from dorsal mHb neurons. Example traces showing typical current steps at −20, 0 and 40 pA in dorsal mHb neurons from wild-type and mutant *Tcf7l2* rats. **j**, Input–output curve in dorsal mHb neurons from wild-type and mutant *Tcf7l2* ($n = 16$ cells from 5 rats) rats. **k**, The frequency at different steps of positive current used to calculate the slope of the input-output curve from ventral mHb neurons. **l**, Input–output curve in ventral mHb neurons from wild-type and mutant *Tcf7l2* ($n = 16$ cells, 5 rats) rats. **m**, Input resistance from mHb neurons from wild-type (13 cells, 4 rats) and mutant (16 cells, 5 rats) *Tcf7l2* rats ($P = 0.1036$, unpaired two-tailed $t$-test). **n**, Afterhyperpolarization in mHb neurons from wild-type (13 cells, 4 rats) and mutant (16 cells, 5 rats) *Tcf7l2* rats; $P = 0.3043$, unpaired two-tailed $t$-test. **o**, Sag current in mHb neurons wild-type (13 cells, 4 rats) and mutant (17 cells, 5 rats) *Tcf7l2* rats ($P = 0.1386$, unpaired two-tailed $t$-test). **p**, Total distance travelled by drug-naive wild-type ($n = 6$) and mutant ($n = 5$) *Tcf7l2* rats during a 60 min session. **q**, Total distance travelled by wild-type ($n = 6$) and mutant ($n = 5$) *Tcf7l2* rats after daily injections of saline or nicotine (0.4 mg $\mathrm{kg}^{-1}$) (15 min pre-treatment time). **r**, Responses to the training dose of nicotine (0.03 mg $\mathrm{kg}^{-1}$ per infusion) were assessed in a group of wild-type ($n = 9$) and mutant ($n = 11$) *Tcf7l2* rats on days 1 and 35 of access. Nicotine responses were similar between the wild-type and mutant *Tcf7l2* rats on day 1 of access, but mutant *Tcf7l2* rats escalated their intake such that their responses were higher on day 35 compared with wild-type *Tcf7l2* rats, and compared with their own intake on day 1 ($F_{1,18} = 30.8$, $^{****}P < 0.0001$, interaction effect between 'genotype' and 'session' in two-way ANOVA). Box plots show minimum–maximum range. Data are mean ± s.e.m.



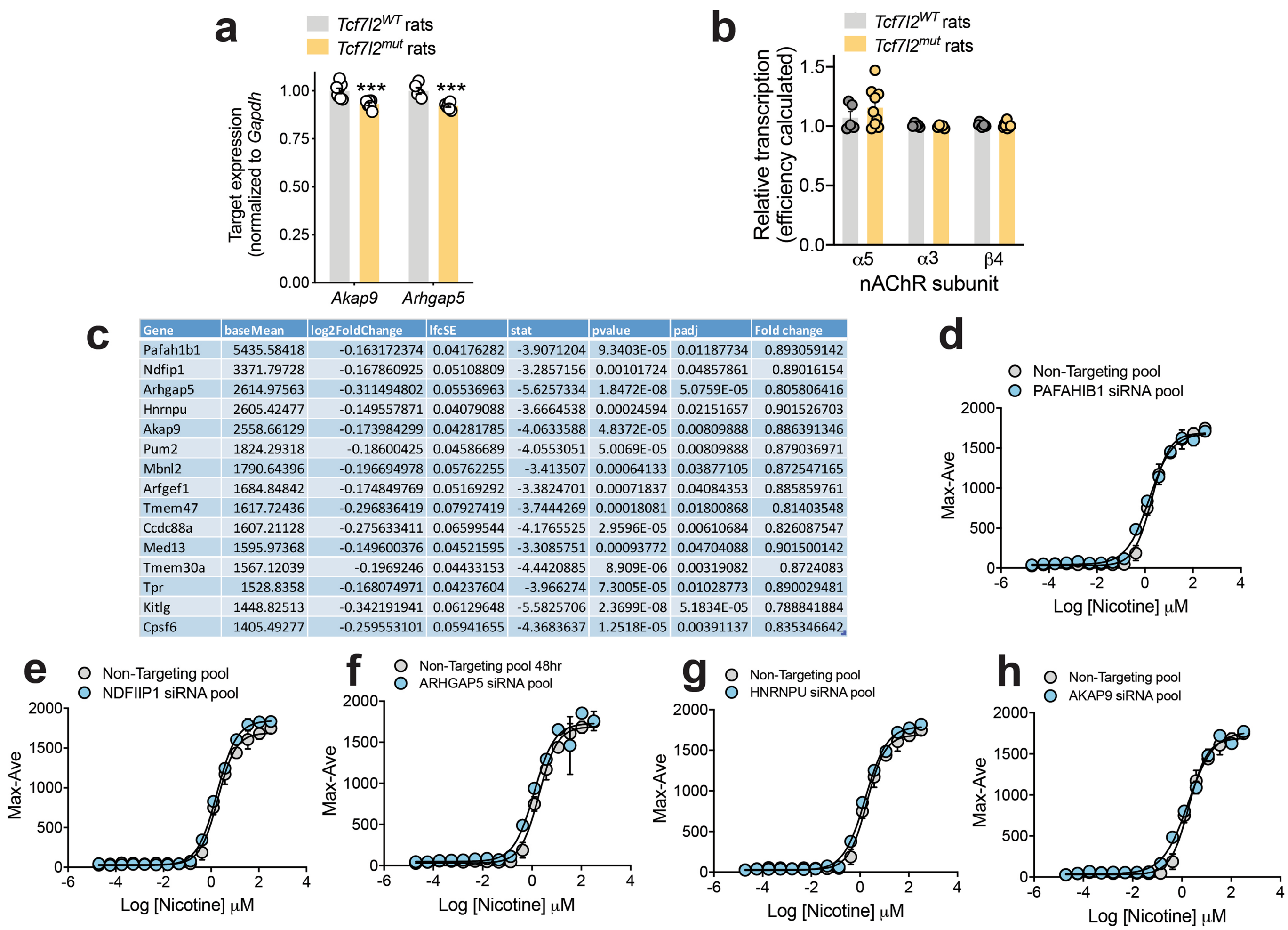
| sgRNA | sgRNA Sequence | Direction | Predicted Off Targets (1 MM-4 MM) | MM not in seed region | Present in all variants |
|---|---|---|---|---|---|
| 1 | GTGTACCCAATCACGACAGG AGG | Forward | 0 - 0 - 1 - 0 - 28 | 0 - 0 - 0 - 0 - 0 | Y |
| 2 | CTGAATCCTCCTGTCGTGAT TGG | Reverse | 0 - 0 - 0 - 1 - 63 | 0 - 0 - 0 - 0 - 0 | Y |
| 3 | TCAGCTGGTAAGTGCGGAGG TGG | Reverse | 0 - 0 - 0 - 9 - 107 | 0 - 0 - 0 - 0 - 0 | Y |
| 4 | AAGTGCGGAGGTGGATTTCC CGG | Reverse | 0 - 0 - 0 - 9 - 64 | 0 - 0 - 0 - 3 - 4 | Y |
| 5 | TCTACGTCAGCTGGTAAGTG CGG | Reverse | 0 - 1 - 0 - 6 - 52 | 0 - 0 - 0 - 0 - 1 | Y |

**Extended Data Fig. 2** | See next page for caption.

**Extended Data Fig. 2 | CRISPR cleavage of *Tcf7l2*.** **a**, Exon diagram of mouse *Tcf7l2* with the two pertinent domains highlighted and the sgRNA targeting locus. **b**, Bioinformatic comparison of the five different sgRNAs tested against *Tcf7l2*. MM, mismatches. **c**, Genomic cleavage percentage in mouse N2a cells of the five sgRNAs targeted against *Tcf7l2*. Data represent $n = 3$ biologically independent samples. **d**, T7-endonuclease-based assay illustrating intact PCR and cleaved bands of *Tcf7l2* via CRISPR gene editing. Observations are from a single experiment. **e**, tdTomato expression in N2a cells 48 h after transduction of the AAV carrying sgRNA against *Tcf7l2* (AAV-sgRNA-Tcf7l2). Data are representative of three biologically independent samples. **f**, Relative expression of *Tcf7l2* transcripts in the N2a cells transfected with AAV-sgRNA-Tcf7l2, AAV-sgRNA-eGFP and AAV-CMV-spCas9 (Vector Biolabs) (***$P < 0.001$, unpaired two-tailed $t$-test). Data represent $n = 5$ biologically independent samples for each gRNA. **g**, Relative mRNA expression of habenular *Tcf7l2* 6 weeks after viral stereotaxic injections of AAV-sgRNA-Tcf7l2, AAV-sgRNA-eGFP or AAV2-hSYN1-iCre into the mHb of $Rosa26^{LSL\text{-}spCas9\text{-}eGFP}$ mice. Data represent $n = 4$ biologically independent samples for each gRNA. **h**, In vivo estimation of genomic cleavage of habenular *Tcf7l2* 6 weeks after viral stereotaxic injection of AAV-sgRNA-Tcf7l2, AAV-sgRNA-eGFP or AAV2-hSYN1-iCre in $Rosa26^{LSL\text{-}spCas9\text{-}eGFP}$ mice. Genomic cleavage efficiency was estimated by average re-annealed mismatches in a T7 endonuclease assay (***$P < 0.001$, unpaired two-tailed $t$-test). Data represent $n = 3$ biologically independent animals for each gRNA. **i**, Left, representative DAPI-counterstained brain slice showing Cas9-eGFP (green) and *Tcf7l2* gRNA (red) targeted to the mHb of $Rosa26^{LSL\text{-}spCas9\text{-}eGFP}$ mice. Right, whole image of brain slice from which left panels are derived. Representative result from $n = 3$ mice. **j**, Medial habenula from $Rosa26^{LSL\text{-}spCas9\text{-}eGFP}$ mice injected with AAV-sgRNA-eGFP ($n = 6$ independent mice) or AAV-sgRNA-Tcf7l2 ($n = 7$ independent mice) was dissected and DNA amplicons of targeted region of *Tcf7l2* sequenced. Percentage of indels detected in the targeted region of *Tcf7l2* is shown (****$P < 0.0001$, unpaired two-tailed $t$-test). Coronal brain image adapted from the Allen Brain Reference Atlas. **k**, Donut graph showing Cas9-induced modifications to *Tcf7l2* in the mHb of $Rosa26^{LSL\text{-}spCas9\text{-}eGFP}$ mice treated with AAV-sgRNA-Tcf7l2 (percentage of total amplicons sequenced). A total of 13 amplicons ($n = 6$ from AAV-sgRNA-eGFP-treated mice and $n = 7$ from AAV-sgRNA-Tcf7l2-treated mice) were sequenced. Data are mean ± s.e.m.

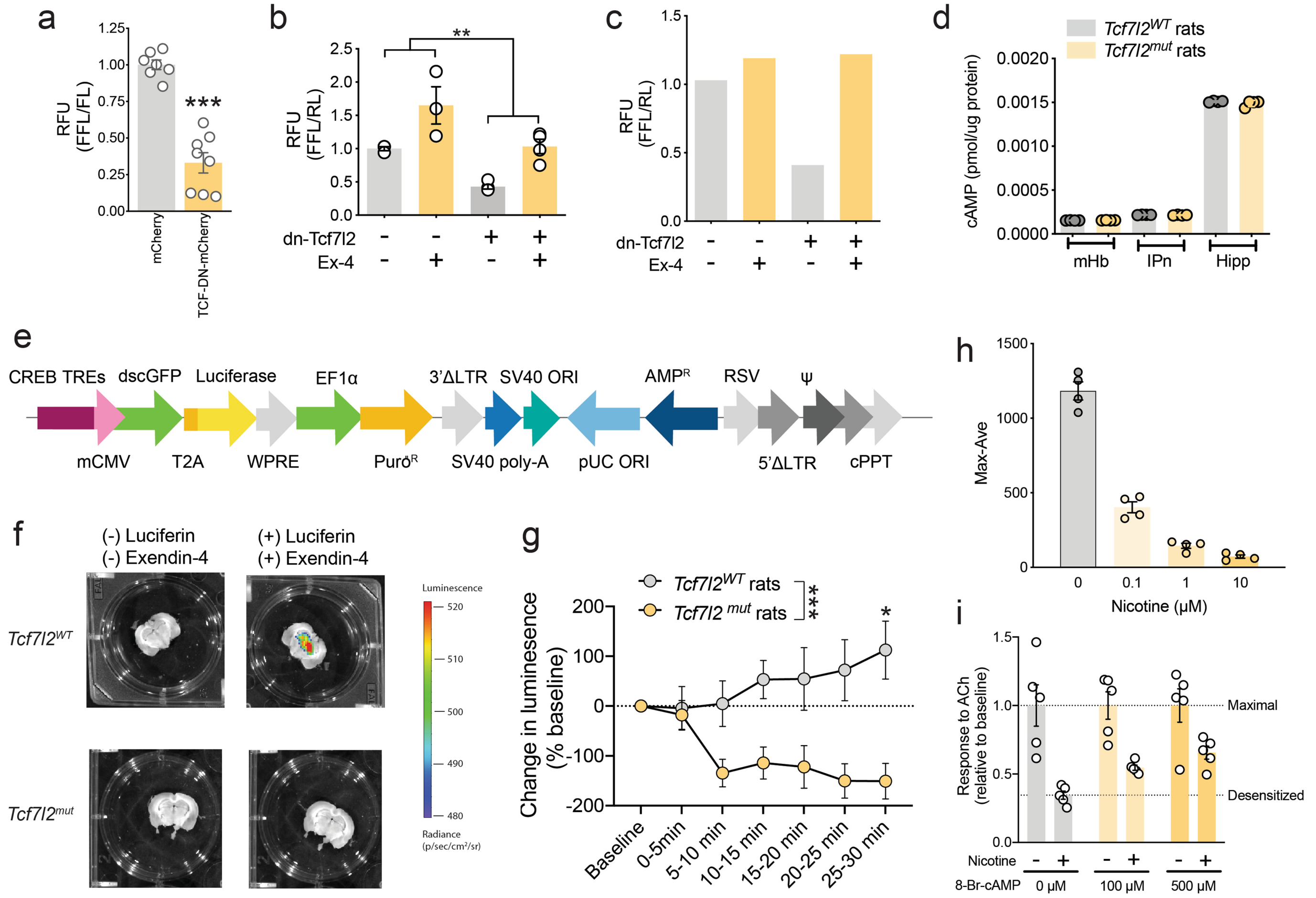


**Extended Data Fig. 3 | Mechanism by which TCF7L2 regulates nAChR function. a**, Effects of intra-mHb infusion of vehicle or EX-4 (12.5–100 ng) on nicotine intake in rats ($n = 10$) ($F_{1.696,\,15.26} = 38.3$, $P < 0.0001$, one-way repeated measures ANOVA; $^{**}P < 0.01$, $^{***}P < 0.001$, Bonferroni's multiple comparisons test). **b**, Effects of intra-mHb infusion of vehicle or EX-4 (100 nM) on the latency to earn the first and second nicotine infusion of a self-administration session in rats ($n = 10$) ($F_{1,\,29} = 311.4$, $P < 0.0001$, main effect of 'infusion number'; $F_{1,\,29} = 125.4$, $P < 0.0001$, main effect of 'EX-4' $F_{1,\,29} = 126.5$, $^{****}P < 0.0001$, interaction effect; two-way ANOVA). **c**, Numbers of rats ($n = 10$ in total) that responded for the first and second nicotine infusion of a self-administration session after intra-mHb infusion of vehicle or EX-4 (12.5–100 ng). **d**, Effects of intra-mHb infusion of vehicle or rDKK1 (100 ng per side) on nicotine intake in rats ($n = 11$) ($P = 0.45$; unpaired two-tailed $t$-test). **e**, Effects of intra-mHb infusion of vehicle or XAV939 (12.5 ng per side) on nicotine intake in rats ($n = 10$) ($P = 0.29$; unpaired two-tailed $t$-test). **f**, Effects of intra-mHb infusion of vehicle or insulin (12.5 ng per side) on nicotine intake in rats ($n = 13$) ($P = 0.29$; unpaired two-tailed $t$-test). **g**–**i**, LiCl (**g**) and constitutively active-β-catenin (CA-β-catenin) (**h**), but not nicotine (**i**), increased GFP relative to mCherry expression in PC12 cells transfected with the 7xTcf-eGFP//SV40-mCherry (7TGC) TCF7L2 reporter. Data reflect results from two independent experiments. **j**, Levels of β-catenin phosphorylated at serine residue 675 or 552 in rat PC12 cells after treatment with forskolin, WNT3A or nicotine. Data reflect results from two independent experiments. **k**, *LacZ* expression in the mHb of BAT-GAL β-galactosidase reporter mice after nicotine injection. Data reflect results from two independent animals in each group. **l**, Expression levels of TCF7L2 (about 69 kDa) in the habenula were measured by western blotting in rats that responded for intravenous nicotine infusions (0.18 mg kg$^{-1}$ per infusion; $n = 12$) or food rewards ($n = 12$). Each lane contains pooled tissues from $n = 3$ animals. Experiment was performed on a single occasion. For uncropped gel image, see Supplementary Fig. 1. **m**, siRNA-mediated knockdown of TCF7L2 attenuated intracellular calcium transients induced by nicotine (20 nM–320 μM) in HEK293T cells heterologously expressing α5α4β2 nAChRs (two-way repeated measures ANOVA; 'siRNA': $F_{1,\,4} = 63.38$, $P < 0.005$; 'nicotine': $F_{15,\,60} = 1388$, $P < 0.0001$; 'siRNA × nicotine': $F_{15,\,60} = 20.89$, $^{***}P < 0.0001$; Bonferroni post hoc test after interaction effect in two-way ANOVA). Representative result from three experiments. **n**, $^{86}$Rb$^{+}$ efflux from synaptosomes generated from IPn tissues derived from wild-type ($n = 6$) and mutant ($n = 6$) *Tcf7l2* rats ($F_{1,39} = 4.267$, $^{*}P = 0.045$; extra sum-of-squares $F$ test). Shift in half-maximal effective concentration ($EC_{50}$) value between genotypes using comparison of fits in a nonlinear fit model. **o**, Pharmacologically isolated nAChR currents (normalized) evoked by nicotine (0.1 Hz application) in mHb neurons from wild-type rats ($n = 3$ cells from 1 rat) were rapidly and completely blocked by bath application of mecamylamine (10 μM). $^{*}F_{1.335,\,2.670} = 332.5$; $P < 0.001$; one-way repeated measures ANOVA. **p**, Baseline nAChR currents in mHb neurons from wild-type ($n = 13$ cells, 4 rats) and mutant ($n = 15$ cells, 4 rats) *Tcf7l2*rats. $P = 0.8180$, unpaired two-tailed $t$-test. **q**, nAChR current decay time after nicotine stimulation (0.1 Hz) before and after nicotine (1 Hz)-induced desensitization in mHb neurons from wild-type ($n = 13$ cells, 4 rats) and mutant ($n = 9$ cells, 4 rats) *Tcf7l2* rats; $P = 0.7133$, unpaired two-tailed $t$-test. **r**, Slope of nAChR current decay after nicotine stimulation (0.1 Hz) before and after nicotine (1 Hz)-induced desensitization in mHb neurons from wild-type ($n = 13$ cells, 4 rats) and mutant ($n = 9$ cells, 4 rats) *Tcf7l2* rats; $P = 0.645$, unpaired two-tailed $t$-test. Data are mean ± s.e.m.

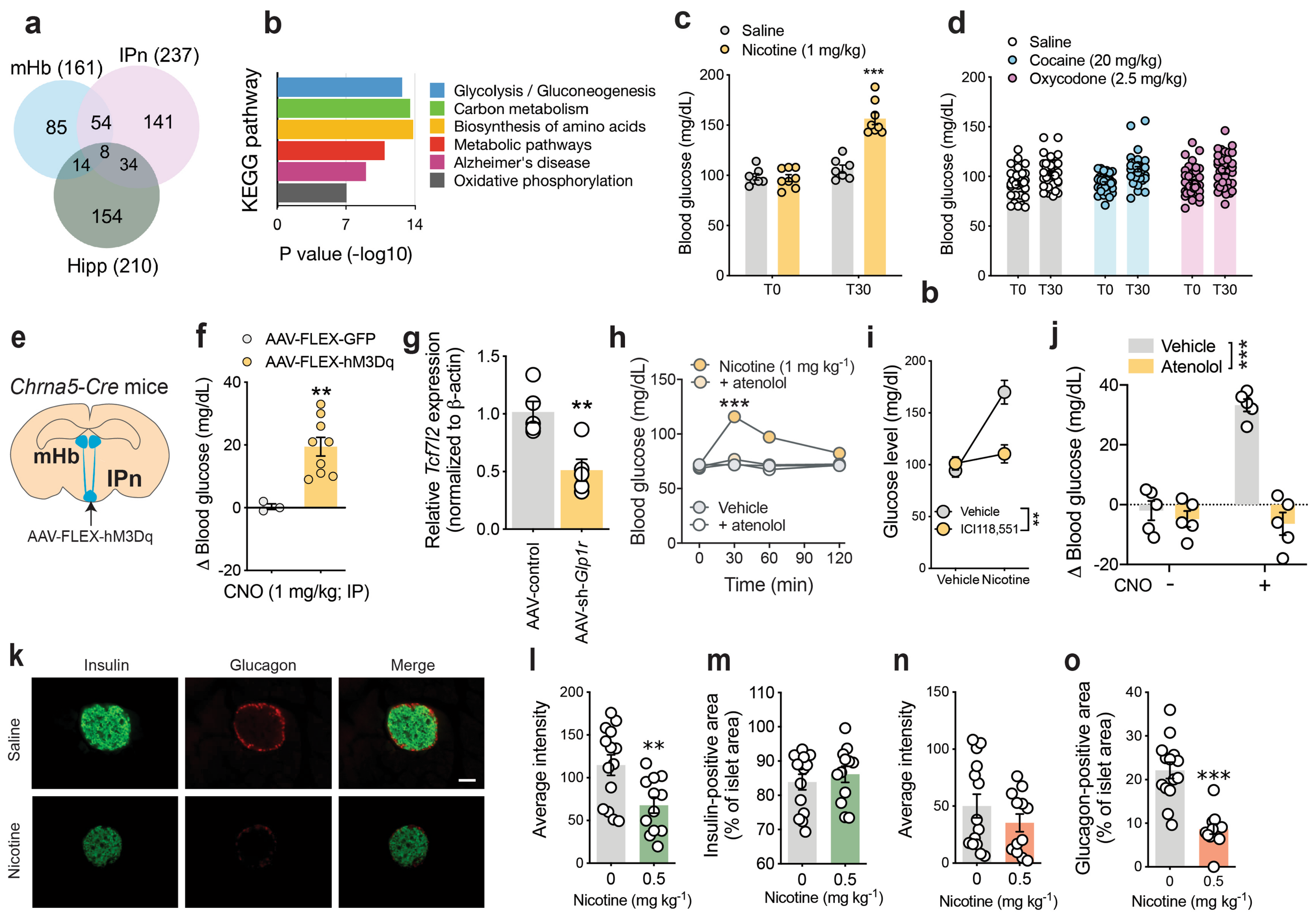


| Gene | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj | Fold change |
|---|---|---|---|---|---|---|---|
| Pafah1b1 | 5435.58418 | -0.163172374 | 0.04176282 | -3.9071204 | 9.3403E-05 | 0.01187734 | 0.893059142 |
| Ndfip1 | 3371.79728 | -0.167860925 | 0.05108809 | -3.2857156 | 0.00101724 | 0.04857861 | 0.89016154 |
| Arhgap5 | 2614.97563 | -0.311494802 | 0.05536963 | -5.6257334 | 1.8472E-08 | 5.0759E-05 | 0.805806416 |
| Hnrnpu | 2605.42477 | -0.149557871 | 0.04079088 | -3.6664538 | 0.00024594 | 0.02151657 | 0.901526703 |
| Akap9 | 2558.66129 | -0.173984299 | 0.04281785 | -4.0633588 | 4.8372E-05 | 0.00809888 | 0.886391346 |
| Pum2 | 1824.29318 | -0.18600425 | 0.04586689 | -4.0553051 | 5.0069E-05 | 0.00809888 | 0.879036971 |
| Mbnl2 | 1790.64396 | -0.196694978 | 0.05762255 | -3.413507 | 0.00064133 | 0.03877105 | 0.872547165 |
| Arfgef1 | 1684.84842 | -0.174849769 | 0.05169292 | -3.3824701 | 0.00071837 | 0.04084353 | 0.885859761 |
| Tmem47 | 1617.72436 | -0.296836419 | 0.07927419 | -3.7444269 | 0.00018081 | 0.01800868 | 0.81403548 |
| Ccdc88a | 1607.21128 | -0.275633411 | 0.06599544 | -4.1765525 | 2.9596E-05 | 0.00610684 | 0.826087547 |
| Med13 | 1595.97368 | -0.149600376 | 0.04521595 | -3.3085751 | 0.00093772 | 0.04704088 | 0.901500142 |
| Tmem30a | 1567.12039 | -0.1969246 | 0.04433153 | -4.4420885 | 8.909E-06 | 0.00319082 | 0.8724083 |
| Tpr | 1528.8358 | -0.168074971 | 0.04237604 | -3.966274 | 7.3005E-05 | 0.01028773 | 0.890029481 |
| Kitlg | 1448.82513 | -0.342191941 | 0.06129648 | -5.5825706 | 2.3699E-08 | 5.1834E-05 | 0.788841884 |
| Cpsf6 | 1405.49277 | -0.259553101 | 0.05941655 | -4.3683637 | 1.2518E-05 | 0.00391137 | 0.835346642 |

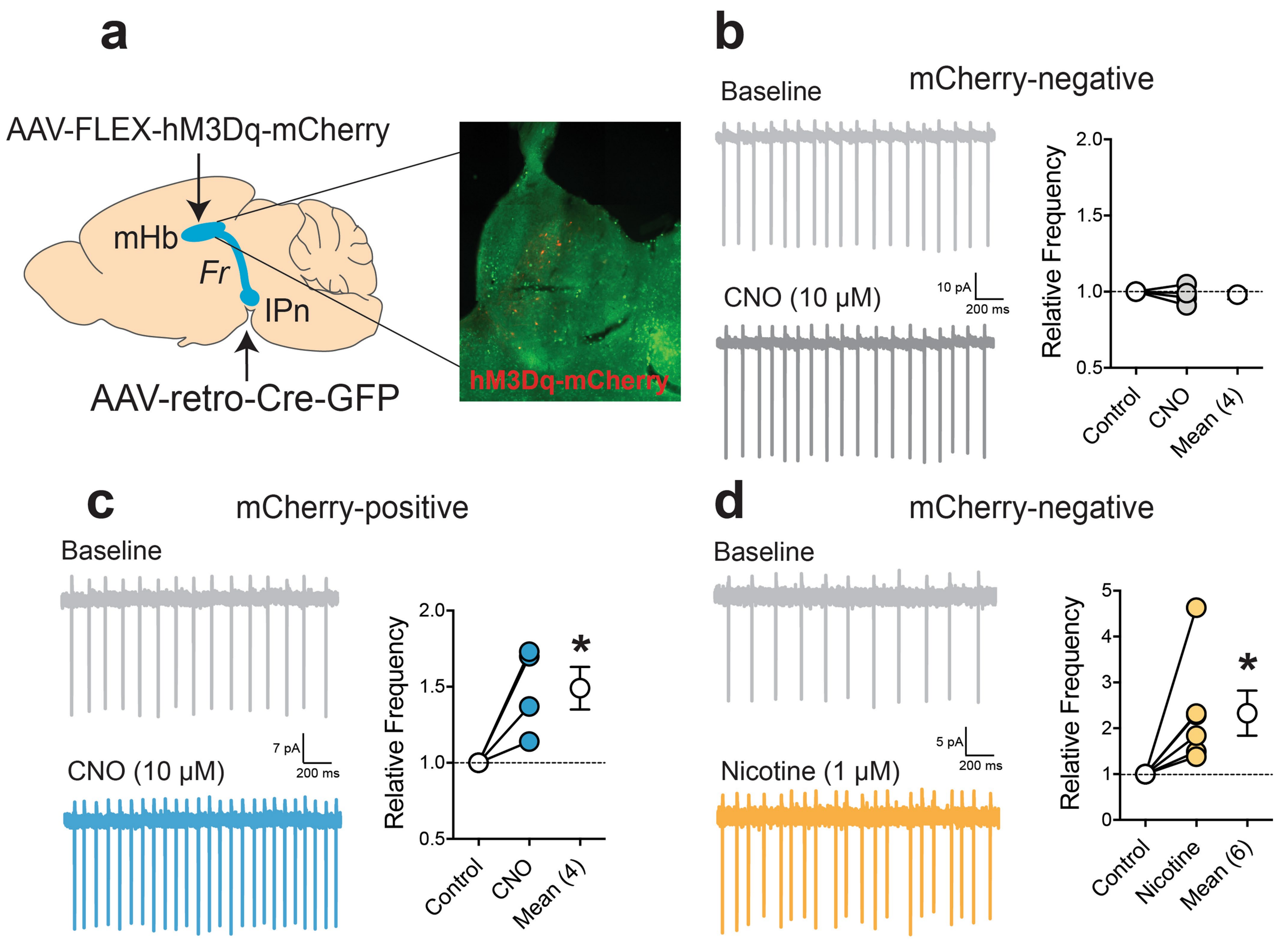
**Extended Data Fig. 4 | Genes regulated by TCF7L2 in the mHb.** **a**, qPCR analysis *Akap9* transcript levels in the mHb of wild-type ($n = 8$) and mutant ($n = 7$) *Tcf7l2* rats; and *Arhgap5* transcript levels in the mHb of wild-type ($n = 5$) and mutant ($n = 7$) *Tcf7l2* rats (***$P < 0.001$, unpaired two-tailed *t*-test). **b**, qPCR analysis of α5, α3 and β4 nAChR subunit expression in the mHb of wild-type ($n = 5$) and mutant ($n = 8$) *Tcf7l2* rats. **c**, Top 15 most abundantly expressed genes in the mHb of wild-type *Tcf7l2* rats that show differential downregulation in the mHb of mutant *Tcf7l2* rats ($n = 9$ rats per genotype). Genes are organized in descending order according to the baseline expression levels in wild-type rats. The $\log_2$-transformed fold change of downregulated gene expression in the mHb of mutant compared with wild-type *Tcf7l2* rats is shown. BaseMean, mean of normalized counts for all samples; lfcSE, s.e.m. of $\log_2$-transformed fold change; stat, Wald chi-squared test of normalized counts for gene transcript in mutant versus wild-type *Tcf7l2* rats; pval, uncorrected Wald test *P* value; padj, *P* value adjusted for multiple testing using Benjamini–Hochberg to estimate the false discovery rate. **d**–**h**, Knockdown of *PAFAHIB1* (**d**), *NDFIIP1* (**e**), *ARHGAP5* (**f**), *HNRNPU* (**g**) or *AKAP9* (**h**) mRNA transcripts using a pool of validated siRNAs had no effects on nicotine-stimulated increases in $[Ca^{2+}]_i$ in human HEK293T cells stably expressing α5α4β2 nAChRs. Data represent $n = 3$ biologically independent samples. Data are mean ± s.e.m.



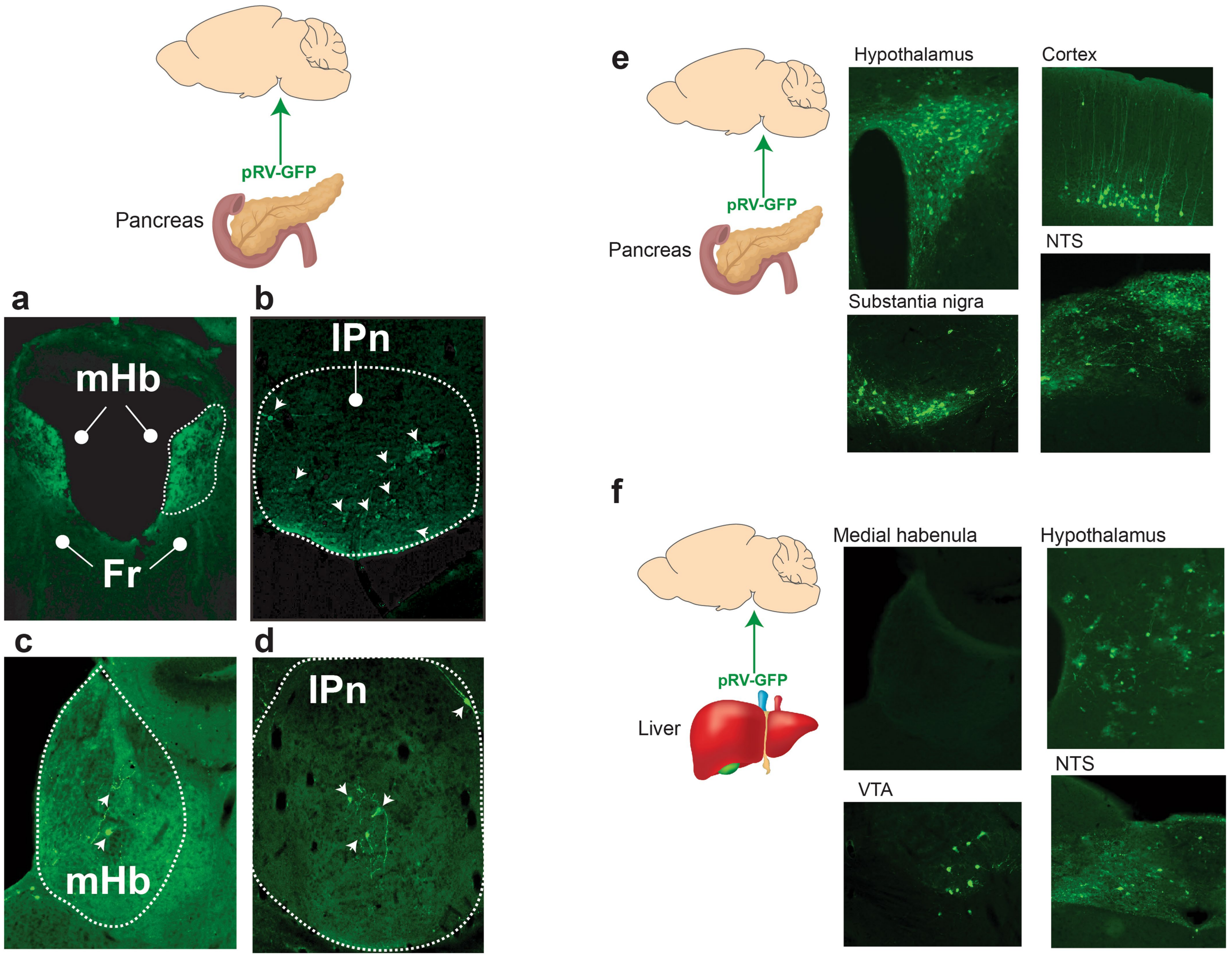
**Extended Data Fig 5 | TCF7L2 regulates cAMP signalling in PC12 cells. a**, Expression of dominant-negative *Tcf7l2* (dnTcf7l2) in PC12 cells reduced the activity of a cAMP-responsive luciferase reporter (EVX-1 luciferase). $***P < 0.001$, unpaired two-tailed *t*-test. Data represent biologically independent samples from cells transfected with mCherry ($n = 7$) or dnTcf7l2 ($n = 8$). FFL, firefly luciferase; RFU, relative fluorescent units; RL, *Renilla* luciferase. **b**, dnTcf7l2 reduced baseline and EX-4-induced increases in a cAMP-responsive reporter assay in PC12 cells ($F_{1,10} = 19.16$, $**P < 0.0014$, main effect of 'dnTcf7l2'; $F_{1,10} = 21.31$, $P = 0.001$, main effect of 'EX-4'; $F_{1,10} = 0.027$, $P = 0.87$, 'dnTcf7l2 × EX-4' interaction; two-way ANOVA). Data represent biologically independent samples from control cells ($n = 3$ samples), control cells treated with EX-4 ($n = 3$ samples), cells transfected dnTcf7l2 ($n = 4$ samples) and cells transfected dnTcf7l2 and treated with EX-4 ($n = 4$ samples). **c**, dnTcf7l2 reduced baseline and EX-4-evoked increases in EVX-1-luciferase in INS-1 cells, an immortalized rat pancreatic β cell line that constitutively expresses GLP-1 receptors. Data represent results from a single experiment. **d**, cAMP content of mHb, IPn and hippocampus were analysed in tissues from wild-type and mutant *Tcf7l2* rats. Each sample contained mHb tissue from 3 rats, and data are from 4 independent samples were analysed for a total of 12 rats per genotype. **e**, Vector map for the pGF-CREB-mCMV-dscGFP-P2A-luciferase (CREB reporter) lentivirus. **f**, Brain slices containing the mHb from wild-type and mutant *Tcf7l2* rats injected with CREB reporter lentivirus into mHb and injected with luciferin just before brain collection. **g**, EX-4 increased luciferase activity in the mHb of wild-type ($n = 3$) but not mutant ($n = 4$) *Tcf7l2* rats ($F_{1,11} = 9.398$, $P = 0.0107$, main effect of 'genotype'; $F_{6,66} = 7.945$, $***P < 0.0001$, interaction effect between 'genotype' and 'EX-4'; two-way repeated-measures ANOVA). **h**, Pre-incubation (30 min) of HEK293T cells stably expressing α5α4β2 nAChRs with nicotine (0.1–10 μM) decreased the ability of acetylcholine (0.1 mM) to stimulate increases in $[Ca^{2+}]_i$ ($F_{3,12} = 188.1$, $P < 0.0001$, main effect of 'nicotine' on one-way ANOVA). Data represent $n = 4$ independent experiment. **i**, 8-Br-cAMP (100–500 μM) attenuated the inhibitory effects of nicotine (0.1 μM) preincubation (30 min) on acetylcholine (0.1 mM) evoked in increases in $[Ca^{2+}]_i$ in α5α4β2 nAChR HEK293T cells ($F_{1,24} = 41.20$, $P < 0.0001$, main effect of 'cAMP' in two-way ANOVA). Data represent $n = 5$ independent experiments. Data are mean ± s.e.m.



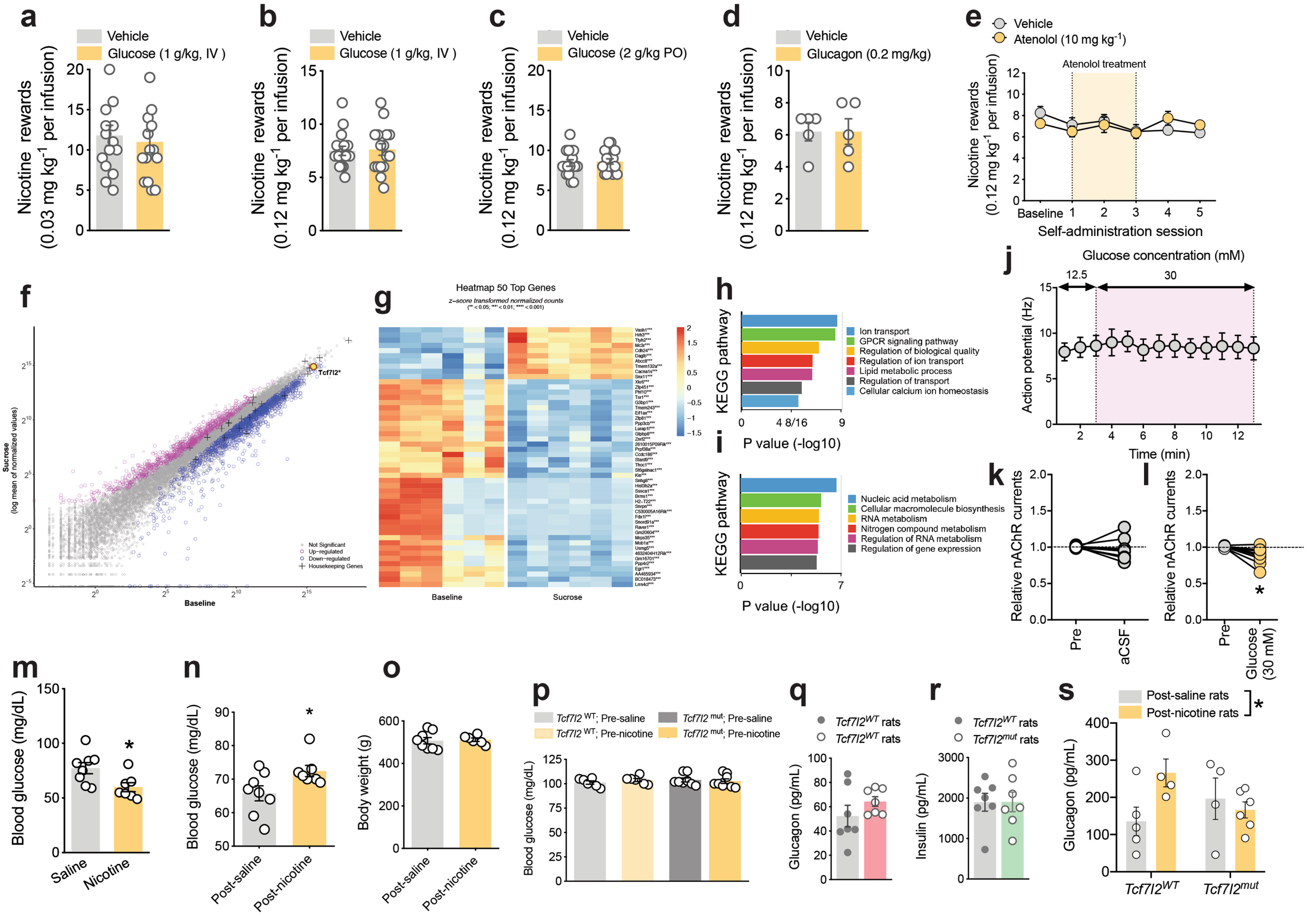
**Extended Data Fig. 6 | Hyperglycaemic actions of nicotine.** **a**, Venn diagram of differentially upregulated genes in the hippocampus, mHb and IPn of mutant *Tcf7l2* rats compared with wild-type rats. **b**, KEGG analysis of differentially upregulated genes identified processes relevant to glucose metabolism as those most likely to be perturbed in the mHb of mutant *Tcf7l2* rats ($n = 9$) compared with wild-type rats ($n = 9$). $P$ values determined by Fisher exact test. **c**, Blood glucose was measured before (T0) and 30 min after (T30) rats ($n = 7$) were injected with saline or nicotine (1 mg kg$^{-1}$) ($F_{1,13} = 52.3$, $^{***}P < 0.0001$, interaction effect between 'nicotine' and 'time'; two-way repeated-measures ANOVA). **d**, Oxycodone (2.5 mg kg$^{-1}$) or cocaine (20 mg kg$^{-1}$) injection had no effects on blood glucose in rats ($n = 36$). **e**, *Chrna5-cre* mice were injected into the IPn with FLEX-GFP ($n = 3$) or FLEX-hM3Dq-mCherry ($n = 9$). Image adapted from the Allen Brain Reference Atlas. **f**, Blood glucose was measured in both groups of mice before and 30 min after injection of CNO (1 mg kg$^{-1}$); $^{**}P < 0.0051$, unpaired two-tailed $t$-test. **g**, *Tcf7l2* mRNA expression was reduced in the mHb of rats after shRNA-mediated knockdown of *Glpr1* transcript expression. $^{**}P < 0.0051$, unpaired two-tailed $t$-test. **h**, Atenolol abolished the hyperglycaemic response to experimenter-administered nicotine injection (1 mg kg$^{-1}$) in rats ($n = 8$) ('atenolol', $F_{3,25} = 43.54$, $P < 0.0001$; 'time', $F_{2.406,60.15} = 48.69$, $P < 0.0001$; 'atenolol × time' interaction, $F_{9,75} = 26.88$; $^{***}P < 0.0001$; two-way ANOVA). **i**, ICI118,551 abolished the increases in blood glucose induced by experimenter-administered nicotine injection (1 mg kg$^{-1}$) in rats ($n = 8$) ('nicotine', $F_{1,7} = 50.83$, $P = 0.002$; 'ICI118,551', $F_{1,7} = 13.17$, $P = 0.0084$; 'ICI118,551 × nicotine' interaction, $F_{1,7} = 27.75$, $^{**}P = 0.0012$; two-way repeated-measured ANOVA). **j**, Atenolol abolished the increases in blood glucose induced by CNO (3 mg kg$^{-1}$) in rats ($n = 8$) expressing FLEX-hM3Dq in the mHb–IPn circuit ('CNO', $F_{1,8} = 213.0$, $P < 0.0001$; 'atenolol', $F_{1,8} = 27.00$, $P = 0.0008$; 'CNO × atenolol' interaction, $F_{1,8} = 255.5$, $^{***}P < 0.0001$; two-way repeated-measures ANOVA). **k**, Immunostaining for insulin (left), glucagon (middle) and their overlap (right) in mice treated acutely with saline (top; $n = 3$) or nicotine (0.5 mg kg$^{-1}$; bottom; $n = 3$). **l–o**, Quantification of insulin intensity (**l**), insulin relative area (**m**), glucagon intensity (**n**) and glucagon relative area (**o**) in pancreatic islets from the saline-treated ($n = 3$) and nicotine-treated ($n = 3$) mice ($^{**}P = 0.0059$, $^{***}P < 0.001$ $t$-test, unpaired two-tailed $t$-test). Image is representative of results obtained in from three biologically independent animals in each treatment group. Data are mean ± s.e.m.



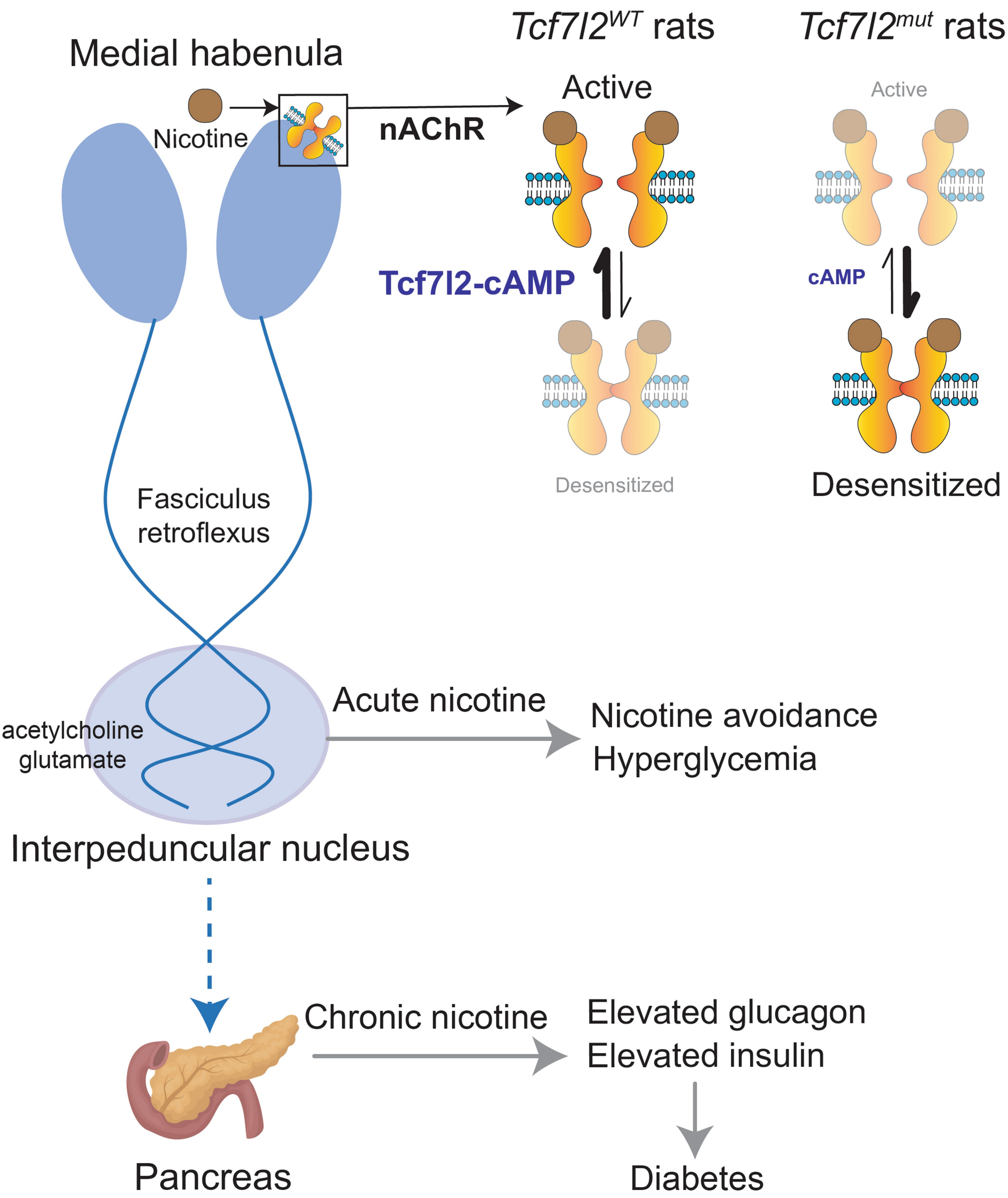
**Extended Data Fig. 7 | Chemogenetic stimulation of the habenula.** **a**, Rats were injected with AAV-retro-Cre into the IPn and FLEX-GFP or FLEX-hM3Dq-mCherry into the mHb. mCherry-positive cells were detected in the mHb, confirming that virus targeting was effective. **b**, CNO (10 μM) had no effects on the relative spike frequency of mCherry-negative cells ($n = 4$ cells, 2 rats). **c**, CNO (10 μM) increased the relative spike frequency of mCherry-positive cells ($n = 4$ cells, 3 rats) ($^*P = 0.0124$, unpaired two-tailed $t$-test). **d**, Nicotine (1 μM) increased the relative spike frequency of mHb neurons by a magnitude similar to that seen in mCherry-positive neurons after CNO treatment ($n = 6$ cells, 3 rats) ($^*P = 0.042$, unpaired two-tailed $t$-test). Data are mean ± s.e.m.



**Extended Data Fig. 8 | pRV mapping of polysynaptic projections from brain to pancreas and liver. a**, **c**, Images of a pRV-GFP-labelled cells (indicated by white arrows) and fibres in the mHb. **b**, **d**, Representative images of pRV-GFP-labelled IPn neurons (indicated by white arrows). **e**, Images of GFP-labelled cells in hypothalamus, cortex, substantia nigra and nucleus of the solitary tract (NTS) after pancreas injection of pRV-GFP. **r**, Images of a GFP-labelled cells in hypothalamus, ventral tegmental area (VTA) and NTS after liver injection of pRV-GFP. Note the absence of GFP-positive cells in the medial habenula. Images are representative of results obtained from three separate experiments. Data are mean ± s.e.m.



**Extended Data Fig. 9 | Consequences of hyperglycaemic actions of nicotine. a**, Effects of glucose (1 mg kg$^{-1}$, intravenous) on nicotine (0.03 mg kg$^{-1}$ per infusion) intake in rats ($n = 15$). **b**, Effects of glucose (1 mg kg$^{-1}$, intravenous) on nicotine (0.12 mg kg$^{-1}$ per infusion) intake in rats ($n = 16$). **c**, Effects of glucose (2 mg kg$^{-1}$, orally) on nicotine (0.12 mg kg$^{-1}$ per infusion) intake in rats ($n = 16$). **d**, Effects of glucagon (0.2 mg kg$^{-1}$, intravenous) on nicotine (0.12 mg kg$^{-1}$ per infusion) intake in rats ($n = 5$). **e**, Atenolol (10 mg kg$^{-1}$) delivered before the self-administration on three consecutive days did not alter nicotine (0.12 mg kg$^{-1}$ per infusion) intake in rats ($n = 8$). **f**, Scatter plots of average TRAP immunoprecipitation samples from sucrose-drinking ($y$ axis; $n = 28$) versus immunoprecipitation samples from sucrose-naive ($x$ axis; $n = 8$) *ChAT*$^{DW167}$ mice representing increased (>0.5 log$_2$-transformed fold change, magenta) or decreased (< −0.5 log$_2$-transformed fold change, blue) levels of transcripts undergoing translation (tissues from $n = 4$ mice were pooled for each sample; 7 samples from sucrose-drinking and 2 samples from sucrose-naive mice were used). Differentially expressed genes were identified by performing a negative binomial test using DESeq2, with default settings. Significant $P$ values were corrected to control the false discovery rate of multiple testing according to the Benjamini–Hochberg procedure at 0.05 threshold and minimum threshold of 0.6 log$_2$-transformed fold change. **g**, Expression levels ($z$-score-transformed normalized counts) of the top 50 genes affected by sucrose consumption in mHb cholinergic neurons. **h**, KEGG analysis of differentially upregulated genes in the mHb of *ChAT*$^{DW167}$ mice described in **f** identified pathways that are likely to be affected in the mHb by sucrose consumption. $P$ values determined by Fisher exact test. **i**, KEGG analysis of differentially downregulated genes in the mHb of *ChAT*$^{DW167}$ mice described in **f** identified pathways that are likely to be affected in the mHb by sucrose consumption. $P$ values determined by Fisher exact test. **j**, The frequency of action potentials in mHb neurons was unaltered by increasing the glucose concentrations in the extracellular solution from 12.5 to 30 mM ($n = 6$ cells from 3 rats). **k**, Maintaining glucose concentration in artificial cerebrospinal fluid (aCSF) in the extracellular solution at 12.5 mM did not alter the magnitude of nicotine-evoked nAChR currents in mHb neurons ($n = 6$ cells from 3 rats). **l**, Increasing the glucose concentration in the extracellular solution from 12.5 to 30 mM decreased the magnitude of nicotine (1 μM) evoked nAChR currents in mHb neurons ($n = 6$ cells from 3 rats) ($^*P < 0.0121$, unpaired two-tailed $t$-test). **m**, Blood glucose levels measured in rats 24 h after their final nicotine (0.12 mg kg$^{-1}$ per infusion; $n = 7$) or saline ($n = 8$) self-administration session ($^*P < 0.0223$, unpaired two-tailed $t$-test). **n**, Blood glucose levels measured in rats 6 weeks after their final nicotine (0.12 mg kg$^{-1}$ per infusion; $n = 7$) or saline ($n = 8$) self-administration session ($^*P < 0.0371$, unpaired two-tailed $t$-test). **o**, Body weights in post-saline ($n = 8$) and post-nicotine rats ($n = 6$) measured 6 weeks after their final self-administration session. **p**, Fasting blood glucose levels in wild-type ($n = 14$ in total) and mutant ($n = 14$ in total) *Tcf7l2* rats measured before chronic saline or nicotine injections commenced. **q**, Circulating levels of glucagon in nicotine-naive wild-type ($n = 7$) and mutant ($n = 7$) *Tcf7l2* rats. **r**, Circulating levels of insulin in nicotine-naive wild-type ($n = 7$) and mutant ($n = 7$) *Tcf7l2* rats. **s**, Circulating glucagon levels in wild-type ($n = 9$ in total) and mutant ($n = 10$ in total) *Tcf7l2* rats measured before chronic saline or nicotine injections ended ($F_{1,15} = 4.606$, $^*P < 0.0486$, interaction effect of 'genotype × nicotine' in two-way ANOVA). Data are mean ± s.e.m.



**Extended Data Fig. 10 | Proposed mechanism by which TCF7L2 regulates the motivational properties of nicotine and its disruptive effects on blood glucose homeostasis.** A representation of a mHb neuron projecting monosynaptically to the IPn (both in blue), via the fasciculus retroflexus, and to the pancreas via a polysynaptic pathway (broken line) is shown. The mHb neurons expresses nAChRs that are activated by nicotine and that undergo nicotine-induced desensitization. In wild-type rats, nAChRs rapidly recover from desensitization by a process involving cAMP signalling. In mutant *Tcf7l2* rats, cAMP signalling is compromised, which results in persistently desensitized nAChRs and diminished sensitivity of mHb neurons to nicotine. When mHb neurons are activated by nicotine, IPn neurons are stimulated by mHb-derived acetylcholine and glutamate. This triggers nicotine avoidance and a hyperglycaemic response, both of which are attenuated in mutant *Tcf7l2* rats. After chronic exposure to the hyperglycaemic actions of nicotine, circulating levels of the pancreas-derived hormones glucagon and insulin are increased, resulting in a diabetes-like disruption of glucose homeostasis. This diabetes-promoting action of nicotine is also attenuated in mutant *Tcf7l2* rats.

https://doi.org/10.1038/s41586-019-1642-0

# Meridional flows in the disk around a young star

Richard Teague[1]*, Jaehan Bae[2] & Edwin A. Bergin[1]

**Protoplanetary disks are known to possess a variety of substructures in the distribution of their millimetre-sized grains, predominantly seen as rings and gaps[1], which are frequently interpreted as arising from the shepherding of large grains by either hidden, still-forming planets within the disk[2] or (magneto-)hydrodynamic instabilities[3]. The velocity structure of the gas offers a unique probe of both the underlying mechanisms driving the evolution of the disk—such as movement of planet-building material from volatile-rich regions to the chemically inert midplane—and the details of the required removal of angular momentum. Here we report radial profiles of the three velocity components of gas in the upper layers of the disk of the young star HD 163296, as traced by emission from $^{12}$CO molecules. These velocities reveal substantial flows from the surface of the disk towards its midplane at the radial locations of gaps that have been argued to be opened by embedded planets[4–7]: these flows bear a striking resemblance to meridional flows, long predicted to occur during the early stages of planet formation[8–12]. In addition, a persistent radial outflow is seen at the outer edge of the disk that is potentially the base of a wind associated with previously detected extended emission[12].**

We use observations of $^{12}$CO $J = 2$–1 emission from HD 163296 to measure the 3D velocity structure of the gas. These data were originally presented as part of the Disk Substructures at High Angular Resolution Project (DSHARP) Atacama Large Millimeter/submillimeter Array (ALMA) Large Program[1,5], which combined previously analysed lower-spatial-resolution data[4,13]. The disk around HD 163296 is known to host multiple rings of large, millimetre-sized grains trapped within regions of gas pressure maxima[6], with a depletion of the gas within the dust gaps[4], highly suggestive of a planetary origin. In the outer disk, at a radius of approximately 260 astronomical units (AU, where 1 AU is the distance between the Earth and the Sun), a local disturbance in the velocity field is probably driven by a massive planet of 2 Jupiter masses ($M_{\rm Jup}$), deeply embedded within the disk[7].

To improve the signal-to-noise ratio and achieve high precision in velocity measurements, we first radially bin the spectral data[6]. As $^{12}$CO $J = 2$–1 is optically thick, the $\tau \approx 1$ surface, where $\tau$ is the optical depth of the emission, traces a region typically 2–4 pressure scale heights above the disk midplane, resulting in asymmetries in the observed emission profiles that must be taken into account when radially binning the data[5,14,15]. This is most clearly seen in a map of the line centre, shown in Fig. 1a, that deviates considerably from the symmetric dipole pattern found in geometrically thin disks. As Keplerian rotation dominates the velocity structure, a parametric emission surface can be inferred that correctly accounts for this projection effect, with the best-fit surface shown as black contours in Fig. 1b, c. This approach finds an emission surface consistent with another technique, which yields a non-parametric surface[6,15]; the latter, however, is highly sensitive to noise in the data and so can only poorly constrain the emission surface in the outer disk.

Azimuthally averaged gas velocities are measured by first splitting the disk into annuli of constant radius using the inferred emission surface. Then, for each radius, we use the azimuthal dependence of the Doppler shift of the line centres due to the projection of the gas velocities, and infer the azimuthally averaged rotational and radial velocities, respectively[6,16] $v_\phi$ and $v_R$. As the $v_\phi$ and $v_R$ components are orthogonal to one another, their projected components have a different azimuthal dependence and can therefore be readily disentangled. Deviations in the line centre at a specific radius relative to the systemic velocity, $v_{\rm LSR}$, is interpreted as a vertical component, $v_Z$. This approach extends the method in refs. [6,16], which considered only $v_\phi$ components. Our improved method recovers the projected components $v_{\phi,{\rm proj}}$ and $v_{R,{\rm proj}}$ absolutely, while determination of $v_{Z,{\rm proj}}$ and the correction into disk-frame velocities, ($v_\phi$, $v_R$, $v_Z$), require a precise measurement of $v_{\rm LSR}$ and the disk inclination, $i$, respectively. The former is well constrained from measurement of the emission surface, $v_{\rm LSR} = 5{,}763 \pm 1$ m s$^{-1}$, while the disk inclination is taken from fits to the rings in continuum[5], which find a disk average $i = 46.7° \pm 0.1°$. The resulting statistical uncertainties on $v_\phi$ and $v_R$ are about 10 m s$^{-1}$ while for $v_Z$ they are about 20 m s$^{-1}$ due to the additional step of measuring the line centre. We estimate that the systematic uncertainties associated with the choice of emission surface are equivalent to between 2 and 3 times the statistical uncertainties. However, as discussed in the Methods section, a different emission surface does not appreciably change the observed flow structure.

To relate the gas velocities to the local sound speed, we use the brightness temperature of the optically thick line emission as a measure of the local gas temperature[5], finding values spanning a range from approximately 90 K in the inner 30 AU to 30 K at 400 AU. This results in local sound speeds ranging between 600 m s$^{-1}$ and 300 m s$^{-1}$. As the pressure-supported Keplerian rotation dominates $v_\phi$, we subtract a baseline model, $v_{\phi,{\rm mod}}$, in order to highlight local deviations in $v_\phi$. Figure 2 shows the resulting velocity structure normalized to the local sound speed in both the ($R$, $\phi$) plane and the ($R$, $Z$) plane. We find striking evidence of a highly dynamic disk. In particular, we find three regions of a collapsing flow (Fig. 2b) with gas rotating at slower and faster velocities either side of the maximum $|v_Z|$ (Fig. 2a), coincident in radius with three previously claimed[4–6] planets and gaps in the dust continuum at 87 AU, 140 AU and 237 AU. In addition, we observe a persistent radially outward flow beyond about 300 AU.

The collapsing regions are bounded by annuli of negative and positive residuals of $v_\phi - v_{\phi,{\rm mod}}$, indicative of a pressure minimum and demonstrating the presence of a substantial gap in the gas surface density[6,16]. Gas is seen to flow towards the gap centres before falling into the region of low gas pressure. Promising candidates for driving such flow structures are meridional flows, which have long been predicted from 3D hydrodynamic simulations at the radii of embedded planets, with their origin well understood[8–11,17]. As seen schematically in Fig. 3, embedded planets will open gaps in the gas by driving material away from the planet at the midplane via Lindblad torques[18]. Locally, the gas pressure drops, causing a decrease in the local gas scale height in order for the disk to maintain vertical hydrostatic equilibrium. In turn, this creates a region of low pressure at the disk surface, at heights as much as 3–4 pressure scale heights above the midplane. Gas from surrounding regions of higher pressure will therefore flow towards the region of lower pressure at the gap centre, before falling towards the midplane. It is this gas near the disk surface that is entrained into the flow towards the gap centre and then downwards to the midplane that we trace with optically thick $^{12}$CO observations.

---

[1]Department of Astronomy, University of Michigan, Ann Arbor, MI, USA. [2]Department of Terrestrial Magnetism, Carnegie Institution for Science, Washington, DC, USA. *e-mail: rteague@umich.edu

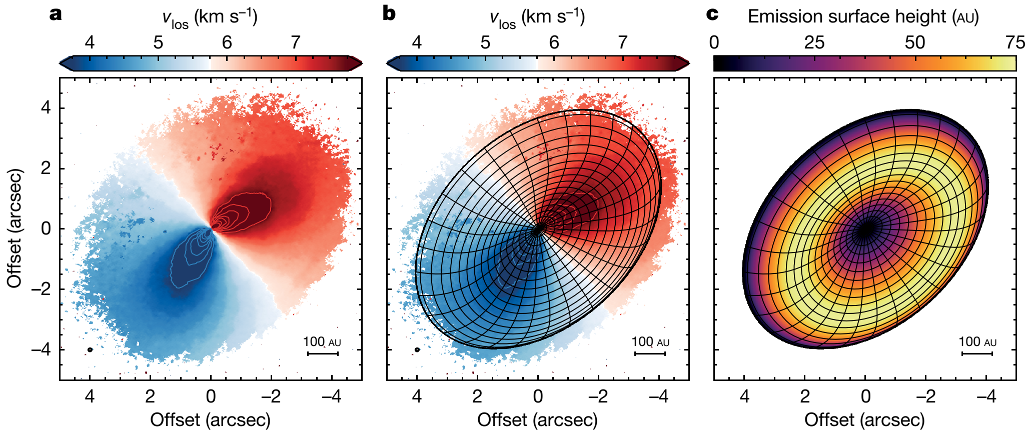


**Fig. 1 | Rotation maps and the inferred 3D geometry of the disk.** **a**, Rotation map (that is, a map of $v_{\rm los}$, the line of sight velocity) of HD 163296 made with the software bettermoments[27]. The filled contours have been clipped to highlight the structure in the disk close to the semi-minor axis. The lined contours are in steps of 0.5 km s$^{-1}$. **b**, As **a** but with an overlay of the inferred emission surface, the height at which the observed emission originates, used to deproject the data. **c**, The inferred height above the midplane of the emission surface. The synthesized beam is shown in the bottom left of **a** and **b** as a filled ellipse.

We run a 3D hydrodynamical simulation of embedded planets in order to provide a qualitative comparison to the observations and demonstrate that $^{12}$CO emission would trace such meridional flows. We inject three planets of masses $0.5M_{\rm Jup}$, $1M_{\rm Jup}$ and $2M_{\rm Jup}$ at 87 AU, 140 AU and 237 AU, respectively, into a disk model that has been shown to recover the disk thermal structure exceptionally well[13,19] (see Methods section). Figure 4 shows the azimuthally averaged density and kinematic structure after 1.44 Myr, a large fraction of the age of the system. The derived velocity structure is in excellent agreement with both the observations and previous simulations, which only contained a single planet-opened gap[8–11,17,20]. This simulation demonstrates that $^{12}$CO emission is able to trace the tops of the meridional flows driven by embedded planets in HD 163296: however, the masses of the inferred planets are model dependent, and more thorough constraints on these are beyond the scope of this work.

Without the direct detection of the embedded planets opening these gaps, other scenarios are possible. For example, zonal flows are known to drive radial deviations in both $v_\phi$ and $v_Z$, allowing also for an efficient cycling of material[3,21,22]. In these simulations, deviations in $v_\phi$ of up to 10% of the local Keplerian rotation are found and similar values in the vertical direction[23]. With the current observations, which lack an absolute scaling of the deviations of $v_\phi$ or $v_Z$, it is impossible to accurately distinguish between scenarios. The use of different molecular tracers offers a potential solution, allowing for the mapping of the flow structures in regions closer to the disk midplane where they are believed to deviate and thus constrain the origin.

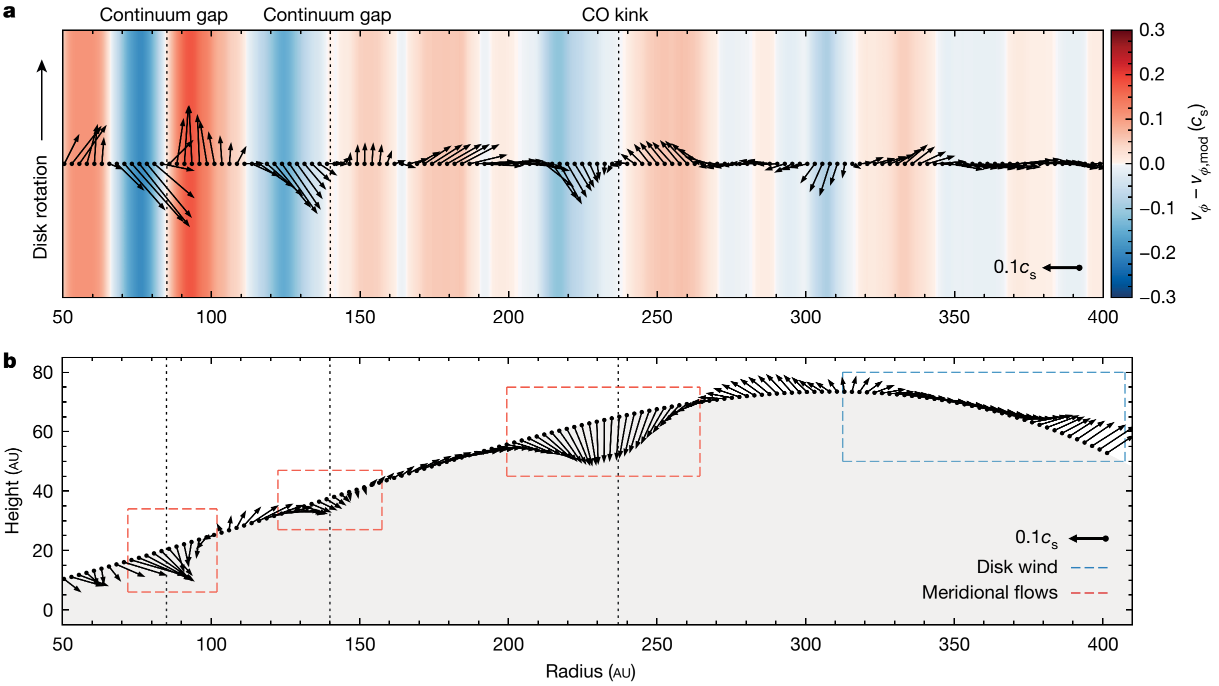


**Fig. 2 | Measured velocity structure of the gas in the disk around HD 163296.** **a**, **b**, Inferred 3D velocity structure in the $(R, \phi)$ plane (**a**), with the disk rotation along the vertical axis, and in the $(R, Z)$ plane (**b**). All velocities are deprojected assuming a source inclination of 46.7° and converted to a fraction of the local sound speed, $c_s$. In both panels, a vector at bottom right shows $0.1c_s$. In **a**, vectors in the positive and negative $y$ direction represent faster- or slower-rotating material, respectively, while the $x$ direction shows the radial flows. The colour background shows the magnitude of the rotation vector ($v_\phi - v_{\phi,\rm mod}$). In **b**, the three locations of meridional circulation are shown in red dashed boxes and the outflow in a blue dashed box. The vertical dotted lines mark the locations of the gaps in the continuum emission[5,28], and the local velocity disturbances, or velocity 'kinks', traced in $^{12}$CO emission[7].

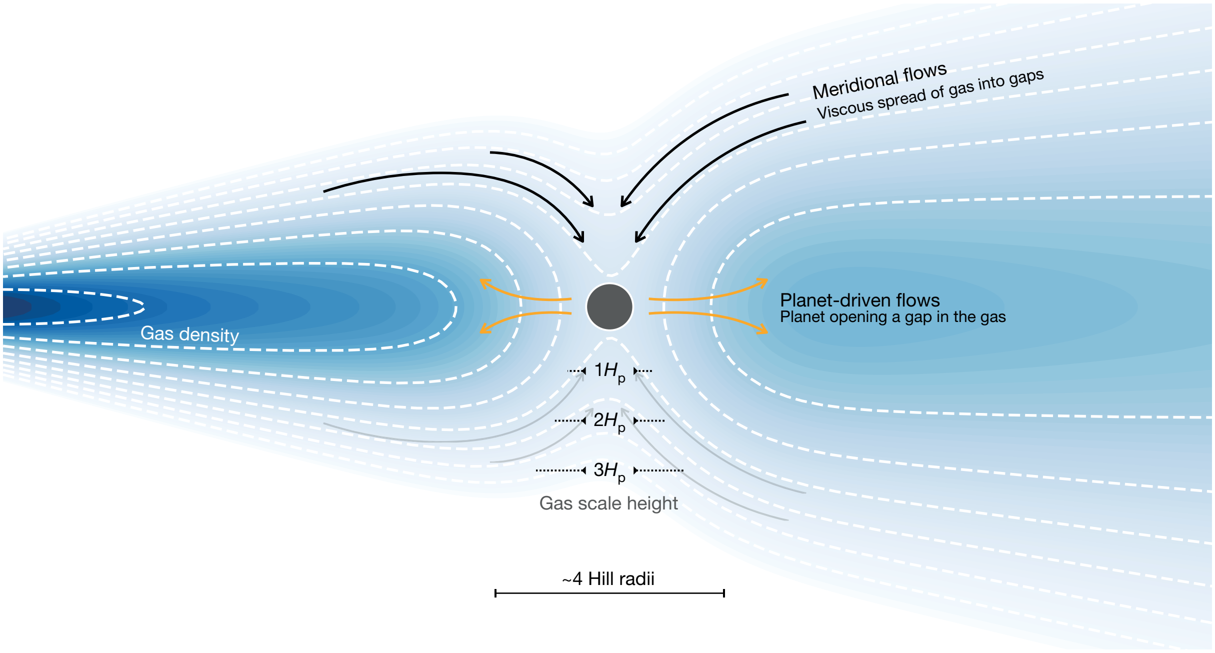


**Fig. 3 | Schematic of the meridional flow.** Cartoon of the dominant flows around an embedded planet, using the approach of ref. [10]. The planet will open a gap in the gas density through Lindblad torques driving the radial flows ('planet-driven flows') shown in orange close to the midplane. The disk will viscously spread towards the gap centre, then fall into the gap via meridional flows, shown with black arrows at higher altitudes. For wide gaps, this will result in circulation around the gap edges. The blue shaded background and white dashed contours show the density profile of the disk. Horizontal black dotted lines show approximate heights in the disk in units of pressure scale heights, $H_\mathrm{p}$.

Regardless of the underlying mechanism driving them, the identification of vertical flows with speeds of about $0.1c_\mathrm{s}$ (where $c_\mathrm{s}$ is the local sound velocity) suggest a rapid cycling of material from the volatile-rich 'molecular layer' of the disk, down towards the considerably cooler midplane, which is substantially more shielded from radiation and therefore more chemically inert. It is this volatile-rich gas that will form the atmospheres of the embedded planet, rather than the midplane material, which is driven away through the Lindblad torques. Therefore, as the carbon-to-oxygen ratio, C/O, is frequently measured in exoplanetary atmospheres and used as a probe of the planet-forming location[24,25], it is the C/O ratio of the in-falling warm molecular regions with which the C/O ratios of exoplanet atmospheres should be compared, rather than the C/O ratio of the disk midplane.

In addition, a radial outflow is found outside about 300 AU, marked with the blue dashed box in Fig. 2b. Although the absolute scaling of $v_Z$ is dependent on the assumed $v_\mathrm{LSR}$, the $v_R$ component is measured

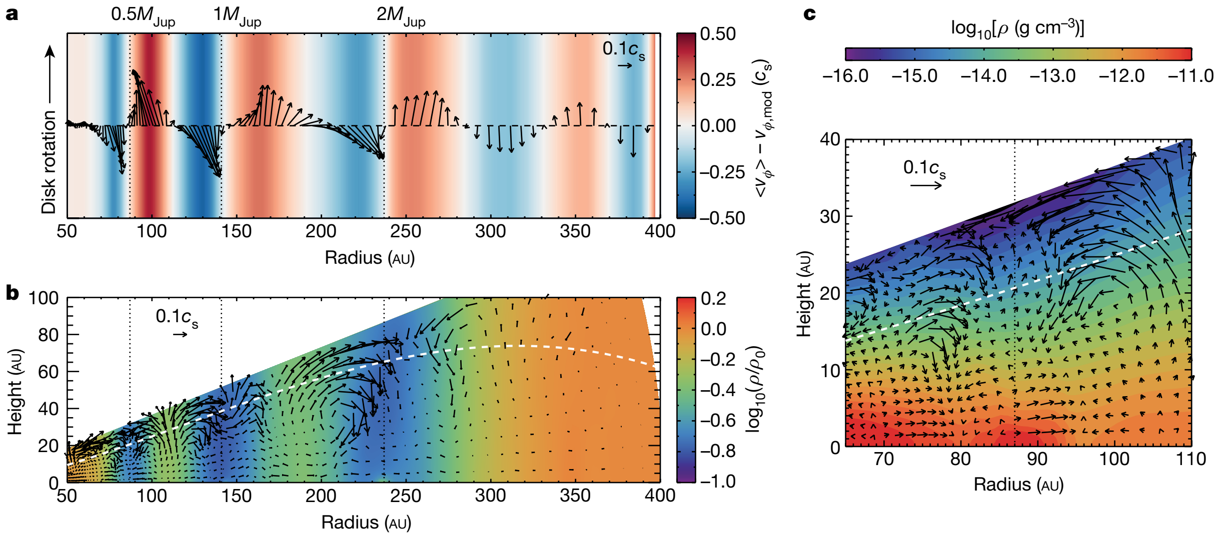


**Fig. 4 | Hydrodynamical simulations of meridional flows.** **a**, **b**, Azimuthally averaged 3D velocity structure in the $(R, \phi)$ plane (**a**) at the $^{12}$CO emission surface and in the $(R, Z)$ plane (**b**), similar to Fig. 2 but from the hydrodynamic simulation using the azimuthally averaged $v_\phi$ velocities. In **b**, the colour contours show the log of the disk gas density ($\rho$) normalized by its initial value in each grid cell ($\rho_0$), and the white dashed curve presents the inferred $^{12}$CO emission surface as in equation (2) (Methods). **c**, A zoomed-in view of the velocity field at the vicinity of the planet at 87 AU. The disk gas density along the $R$–$Z$ slice containing the planet is shown in the background on a logarithmic scale. All the contours and vectors show snapshot values taken at the end of the simulation (1.44 Myr). For visualization purpose only, we present vectors every two radial and meridional grid cells in **a** and **c**, and every three radial and meridional grid cells in **b**.

absolutely so that a true outflow of material is observed, while there is no evidence for such radial motion inward of 300 AU. Viscous spreading of the disk can be ruled out, as this should result in much lower velocities of the order of $<1$ m s$^{-1}$, considerably slower than the measured $v_{\rm flow} \approx 30$ m s$^{-1}$; however, the orientation of the vectors is suggestive that the radial outflow is the base of a disk wind. A large-scale molecular wind, extending over 10″ (about 1,000 AU) and reaching velocities of approximately 20 km s$^{-1}$, has been previously observed in $^{12}$CO around HD 163296 with associated Herbig Haro knots distributed along a jet[12,26]. The large-scale wind kinematics indicate a substantial mass loss rate of $(1.4–18.3) \times 10^{-8}$ solar masses per year. If this large-scale molecular wind is related to the radial outflow we observe, it would suggest that disk winds can provide an efficient mechanism for removing angular momentum from the disk without the need for turbulent viscosity, reconciling the inferred low levels of turbulence[13,19]. As discussed in the Methods section, the details of the flow structure are dependent on the assumed emission surface and therefore any further analysis of the potential launching mechanisms first requires a more robust measure of the emission surface at the disk edges.

The detection of large-scale flows demonstrates that protoplanetary disks are probably still highly dynamic and actively changing their physical and chemical structures. Looking forward, the ability to infer 3D velocities will aid further planet detection at (sub-)millimetre wavelengths and perhaps even aid our understanding of the puzzle of disk angular momentum transport.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1642-0.

Received: 22 March 2019; Accepted: 25 July 2019;
Published online 16 October 2019.

1. Andrews, S. M. et al. The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, sample, calibration, and overview. *Astrophys. J.* **869**, L41 (2018).
2. Zhang, S. et al. The Disk Substructures at High Angular Resolution Project (DSHARP). VII. The planet–disk interactions interpretation. *Astrophys. J.* **869**, L47 (2018).
3. Flock, M. et al. Gaps, rings, and non-axisymmetric structures in protoplanetary disks. From simulations to ALMA observations. *Astron. Astrophys.* **574**, A68 (2015).
4. Isella, A. et al. Ringed structures of the HD 163296 protoplanetary disk revealed by ALMA. *Phys. Rev. Lett.* **117**, 251101 (2016).
5. Isella, A. et al. The Disk Substructures at High Angular Resolution Project (DSHARP). IX. A high-definition study of the HD 163296 planet-forming disk. *Astrophys. J.* **869**, L49 (2018).
6. Teague, R., Bae, J., Bergin, E. A., Birnstiel, T. & Foreman-Mackey, D. A kinematical detection of two embedded Jupiter-mass planets in HD 163296. *Astrophys. J.* **860**, L12 (2018).
7. Pinte, C. et al. Kinematic evidence for an embedded protoplanet in a circumstellar disk. *Astrophys. J.* **860**, L13 (2018).
8. Szulágyi, J., Morbidelli, A., Crida, A. & Masset, F. Accretion of Jupiter-mass planets in the limit of vanishing viscosity. *Astrophys. J.* **782**, 65 (2014).
9. Morbidelli, A. et al. Meridional circulation of gas into gaps opened by giant planets in three-dimensional low-viscosity disks. *Icarus* **232**, 266–270 (2014).
10. Fung, J. & Chiang, E. Gap opening in 3D: single-planet gaps. *Astrophys. J.* **832**, 105 (2016).
11. Dong, R., Liu, S.-Y. & Fung, J. Observational signatures of planets in protoplanetary disks: planet-induced line broadening in gaps. *Astrophys. J.* **870**, 72 (2019).
12. Klaassen, P. D. et al. ALMA detection of the rotating molecular disk wind from the young star HD 163296. *Astron. Astrophys.* **555**, A73 (2013).
13. Flaherty, K. M. et al. Weak turbulence in the HD 163296 protoplanetary disk revealed by ALMA CO observations. *Astrophys. J.* **813**, 99 (2015).
14. Rosenfeld, K. A., Andrews, S. M., Hughes, A. M., Wilner, D. J. & Qi, C. A spatially resolved vertical temperature gradient in the HD 163296 disk. *Astrophys. J.* **774**, 16 (2013).
15. Pinte, C. et al. Direct mapping of the temperature and velocity gradients in discs. Imaging the vertical CO snow line around IM Lupi. *Astron. Astrophys.* **609**, A47 (2018).
16. Teague, R., Bae, J., Birnstiel, T. & Bergin, E. A. Evidence for a vertical dependence on the pressure structure in AS 209. *Astrophys. J.* **868**, 113 (2018).
17. Kley, W., D'Angelo, G. & Henning, T. Three-dimensional simulations of a planet embedded in a protoplanetary disk. *Astrophys. J.* **547**, 457–464 (2001).
18. Lin, D. N. C. & Papaloizou, J. On the tidal interaction between protoplanets and the protoplanetary disk. III – Orbital migration of protoplanets. *Astrophys. J.* **309**, 846–857 (1986).
19. Flaherty, K. M. et al. A Three-dimensional view of turbulence: constraints on turbulent motions in the HD 163296 protoplanetary disk using DCO$^+$. *Astrophys. J.* **843**, 150 (2017).
20. Gressel, O., Nelson, R. P., Turner, N. J. & Ziegler, U. Global hydromagnetic simulations of a planet embedded in a dead zone: gap opening, gas accretion, and formation of a protoplanetary jet. *Astrophys. J.* **779**, 59 (2013).
21. Lyra, W., Johansen, A., Klahr, H. & Piskunov, N. Global magnetohydrodynamical models of turbulence in protoplanetary disks. I. A cylindrical potential on a Cartesian grid and transport of solids. *Astron. Astrophys.* **479**, 883–901 (2008).
22. Johansen, A., Youdin, A. & Klahr, H. Zonal flows and long-lived axisymmetric pressure bumps in magnetorotational turbulence. *Astrophys. J.* **697**, 1269–1289 (2009).
23. Suzuki, T. K. & Inutsuka, S.-i. Magnetohydrodynamic simulations of global accretion disks with vertical magnetic fields. *Astrophys. J.* **784**, 121 (2014).
24. Öberg, K. I., Murray-Clay, R. & Bergin, E. A. The effects of snowlines on C/O in planetary atmospheres. *Astrophys. J.* **743**, L16 (2011).
25. Madhusudhan, N. C/O ratio as a dimension for characterizing exoplanetary atmospheres. *Astrophys. J.* **758**, 36 (2012).
26. Devine, D. et al. A Ly$\alpha$ bright jet from a Herbig AE star. *Astrophys. J.* **542**, L115–L118 (2000).
27. Teague, R. & Foreman-Mackey, D. A robust method to measure centroids of spectral lines. *Res. Not. Am. Astron. Soc.* **2**, 173 (2018).
28. Huang, J. et al. The Disk Substructures at High Angular Resolution Project (DSHARP). II. Characteristics of annular substructures. *Astrophys. J.* **869**, L42 (2018).

# METHODS

**Observations.** We use the publicly available images from the Disk Substructures at High Angular Resolution Program (DSHARP) ALMA Large Program[1,5]. Combining data from three projects, 2013.1.00366.S[13], 2013.1.00601.S[4] and 2016.1.00484.L[1], a spatial resolution of 0.104″ × 0.095″ (corresponding to 10.5 AU × 9.6 AU at the source distance of 101 pc) was achieved, and a spectral resolution of 640 m $s^{-1}$ after Hanning smoothing. More details about the data reduction and imaging procedure can be found in Isella et al.[5].

**Deriving an emission surface.** In order to accurately deproject the observations into annuli of constant radius we must take into account the flared emission surface of the disk. We use two different approaches in order to derive a robust emission surface. The first approach uses the morphology of the map of the line centre. For a geometrically thin disk, the rotation map will have a symmetric dipole pattern. However, with an elevated emission surface, the red- and blue-shifted lobes of the dipole bend away from the disk semi-major axis, allowing for a constraint on the emission surface. We first make a map of the line centres using bettermoments[27] which fits a quadratic curve to the pixel of peak intensity and the two neighbouring pixels. This has the considerable advantage over the more traditionally used intensity-weighted averaged velocities (first moment maps) in that no clipping is required and that the near side of the disk is readily distinguished from the far side, in addition to returning well characterized uncertainties on the line centre. The resulting map of the line centres is shown in Fig. 1a.

Assuming that the kinematics of the disk are dominated by Keplerian rotation, we also fit the rotation map with a Keplerian rotation pattern including the correction for the height above the midplane[14]:

$$v_0 = \sqrt{\frac{GM_{\rm star}R^2}{(R^2+Z^2)^{3/2}}}\,\cos\theta\,\sin(i) \tag{1}$$

where $v_0$ is the projected line centre, $R$ and $Z$ are the midplane radius and the height above the midplane, respectively, $\theta$ is the polar angle of the disk, $i$ is the inclination of the disk and $M_{\rm star}$ is the stellar mass. The emission surface is parameterized as two power-law profiles to capture the flared increasing surface over most of the disk, but also the drop in the outer disk due to the decreasing gas surface density:

$$Z(R) = Z_0 \times (R/1'')^{\varphi} - Z_1 \times (R/1'')^{\psi} \tag{2}$$

where both $Z_0$ and $Z_1$ should be positive, and $\varphi$ and $\psi$ represent the flaring of the emission surface. In addition to the emission surface, we let the source centre and disk inclination and position angle, $\{x_0, y_0, i, {\rm PA}\}$, vary as well as a variable stellar mass, $M_{\rm sun}$, and systemic velocity, $v_{\rm LSR}$, leaving 10 free parameters for the model.

The calculation of the posterior distributions was performed using the Python package eddy[29] assuming a source distance[30] of 101 pc. Only spatially independent pixels, those at least a beam FWHM apart, were considered in the calculation of the likelihood. Flat priors were assumed for all values except the inclination, which has a prior of $i = 46.7° \pm 0.1°$ based on the fits to the continuum observations[5]. 64 walkers were run for 10,000 steps for an initial burn-in period, before an additional 10,000 steps were used to sample the posterior distributions. The resulting posterior distributions were found to be $x_0 = -5^{+2}_{-2}$ mas, $y_0 = 3^{+2}_{-2}$ mas, $i = 46.8°^{+0.1}_{-0.1}$, ${\rm PA} = 312.83°^{+0.02}_{-0.02}$, $M_{\rm star} = 2.022^{+0.004}_{-0.004}\,M_{\rm sun}$, $v_{\rm LSR} = 5.763^{+0.001}_{-0.001}$ km $s^{-1}$ and an emission surface described by $Z_0 = 265^{+3}_{-3}$ mas, $\varphi = 1.29^{+0.04}_{-0.03}$, $Z_1 = 6^{+2}_{-3}$ mas and $\psi = 3.8^{+0.3}_{-0.3}$, where the uncertainties are the 16th to 84th percentiles of the posterior distribution about the median value. We note that these uncertainties represent only the statistical uncertainty on the values; the systematic uncertainties, due to neglecting the pressure correction term in the calculation of the velocity profile, for example, are not included. Fixing $Z_1 = 0$ results in comparable posterior distributions for the geometric parameters and emission surface characterized by $Z_0 = 280^{+2}_{-2}$ mas, but a smaller $\varphi = 0.994^{+0.007}_{-0.007}$ to account for the drop in the emission surface at larger radii, broadly consistent with previous work[5].

The second method follows the method presented in Pinte et al.[15], which allows for a non-parametric surface that can account for a drop of the emission surface into gap regions, as previously demonstrated for $C^{18}O$ emission in this source[6]. This method assumes that the gas is on circular orbits and uses the asymmetry of the emission relative to the major axis of the disk to infer the height of the emission relative to the disk midplane. For this approach we adopt the inclination and position angle found from continuum[5], $i = 46.7°$ and ${\rm PA} = 133.3°$ (note here there is a 180° degeneracy in the PA when modelling continuum emission). Following Teague et al.[6], the raw data points were modelled with a Gaussian process resulting in a smooth emission surface, rather than radially binning the data.

Extended Data Fig. 1 compares the two derived emission surfaces. The grey solid lines show 50 random samples of the parametric surface, while the red dashed line shows the Gaussian process model and the associated $3\sigma$ uncertainties when modelling the raw points shown as grey dots. In the outer disk, $r \gtrsim 3''$, the two surfaces deviate. This is because for the non-parametric method, the fit is dependent on individual pixels, which have a lower signal-to-noise ratio in the outer disk (as demonstrated by the broad scatter in the grey points). Conversely, as the parametric model is less flexible, it is therefore less affected by the noisy outer disk. Using only the data points for $r < 3$ we find a consistent non-parametric emission surface.

**Measuring the velocity structure.** To infer radial profiles of the radial and rotational velocities, $v_R$ and $v_\phi$, respectively, we follow the method outlined in Teague et al. [6,16], using the Python implementation eddy[29]. The disk is first split into annuli of constant radius accounting for the inferred emission surface, each 25 mas wide (roughly 1/4 of the beam FWHM), from which spatially uncorrelated pixels (at least a beam FWHM apart) were randomly selected (the pixel scaling was 13 mas). The line centre of each pixel was assumed to be the sum of the projected component of $v_\phi$, and $v_R$ such that:

$$v_{\rm los}(\phi) = v_{\phi,{\rm proj}}\cos\phi + v_{R,{\rm proj}}\sin\phi \tag{3}$$

where $\phi$ is the deprojected polar angle along the annulus. The best fitting $v_\phi$ and $v_R$ values were those that maximized the likelihood when a Gaussian process was fitted to the shifted and stacked data, which is a robust measure of the alignment of the spectra, as described in Teague et al.[16]. For each annulus, 20 different realizations with different pixels were performed and the posterior distributions combined.

To calculate the projected vertical component of the velocity, $v_{Z,{\rm proj}}$, we fitted a Gaussian to the aligned and stacked spectra using the best-fit $v_{\phi,{\rm proj}}$ and $v_{R,{\rm proj}}$ values. To improve the quality of the Gaussian fit, we resample the stacked spectra to 40 m $s^{-1}$, a factor of 16 higher than the original data. This resampling will not alter the intrinsic line profile and will preserve any systematic effects, such as the spectral response function of the ALMA correlator and the Hanning smoothing. While these systemic effects will bias measurements of the line peak or width, the line centre should remain unaffected. We perform two fits to the data. First, we use the whole spectrum to identify the line centre and a rough line width. Second, with only the line core defined, as where $|v - v_0| \le \Delta V$, in order to remove biases due to the noisy line wings arising from contamination from the far side of the disk (see the appendix in Teague et al.[16]). The measured line centre is then considered the sum of the systemic velocity, $v_{\rm LSR}$, and the projected vertical velocity, $v_{Z,{\rm proj}}$. Without a precise measure of $v_{\rm LSR}$, the inferred $v_Z$ radial profiles are therefore only relative values. In future, observations of an optically thin line where $v_Z$ components would cancel out from either side of the disk may allow for a better determination of the $v_{\rm LSR}$ and constraints on the absolute values of $v_{z,{\rm proj}}$. Finally, the projected velocity components were corrected for their projection using a disk inclination of 46.7°.

We perform this method for both of the emission surfaces we have derived to verify that the deprojection does not substantially change the derived velocity vectors. Furthermore, to check that the inclusion of a $v_R$ term was not changing the inferred $v_\phi$ value, we repeat the procedure fixing $v_{R,{\rm proj}} = 0$ and find no change in the inferred $v_{\phi,{\rm proj}}$. This is expected, because the radial and rotational velocity components are orthogonal and thus $v_R$ components cannot correct for changes in $v_\phi$ and vice versa. We also use the differences between the two methods to estimate the systematic uncertainties on the derived velocities. Across the whole radius of the disk, the average differences between the methods are $1.5\sigma$, $1\sigma$ and $0.2\sigma$ for $v_\phi$, $v_R$ and $v_Z$, respectively. Within the gap regions where the two surfaces are most different, most notably D48 or at 220 AU, these rise to about $3\sigma$ for all three components. Despite the difference in the emission heights, the qualitative flow structures remain, although the wind signature is less well defined than with the smooth parametric surface.

As the perturbations are expected to be of the order of a few per cent of the Keplerian rotation, we must subtract a background profile from $v_\phi$ in order to clearly see them. However, as $v_\phi$ is a combination of the Keplerian rotation (which varies depending on the height of the emission surface) in addition to perturbations from the radial pressure gradient and self-gravity of the disk[14], there is no simple predictive fit to the curve. To act as a model of the background rotation, $v_{\phi,{\rm mod}}$, we fit a 10th-order polynomial to $v_\phi$. We have also tried single and broken power-law functions, but these were unable to reliably capture the overall trend at small and large radii. Extended Data Fig. 5 demonstrates the choice of $v_{\phi,{\rm mod}}$ and how the resulting residuals are affected. We stress that this is for presentation purpose only and, if any quantitative fits were to be made, they should be made directly to $v_\phi$, as in previous works[6,16].

The resulting velocities are plotted in Extended Data Fig. 2 in blue, with the three components in the three lower panels corrected for the projection. The values in red show the results when we fix $v_R = 0$. The locations of the gaps in the continuum[5,28] and previously reported planets[6,7] are annotated in the top row. The results for both types of emission surface are consistent, demonstrating that the results are robust to the choice of emission surface. The inferred flow structures are shown in Fig. 2 for the deprojection using the parametric surface, and in Extended Data Fig. 3 for the deprojection using the non-parametric surface.

To convert the velocities to Mach numbers, we use the peak value from the aligned and deprojected spectra to measure the local gas temperature using the full Planck law and calculate a local sound speed. The derived gas temperatures, consistent with previous analyses[5], and derived sound speeds are shown in Extended Data Fig. 4. This temperature may be underestimated if the emission line is only marginally thick, $\tau \lesssim 5$, however at this temperature this requires a column density of $N(\mathrm{CO}) \lesssim 10^{16}\ \mathrm{cm}^{-2}$, considerably less than found in previous models of CO isotopologue emission in this source[13,19]. Owing to the low spectral resolution of the data and confusion from the far side of the disk, measuring the line width will not provide a reasonable measure of the gas kinetic temperature[16,31].

**Hydrodynamic simulation setup.** We solve the isothermal hydrodynamic continuity equations and the Navier–Stokes equations in 3D spherical $(r, \theta, \phi)$ coordinates using publicly available planet–disk interaction code FARGO3D[32], adopting the orbital advection algorithm[33].

We adopt a parametric disk model similar to the one presented in Flaherty et al.[19], which is shown to reproduce CO isotopologue emission well. The vertically integrated surface density of the disk follows:

$$\Sigma_{\mathrm{gas}}(R) = \frac{M_{\mathrm{gas}}}{2\pi R_{\mathrm{c}}^2}(2-\gamma)\left(\frac{R}{R_{\mathrm{c}}}\right)^{-\gamma} \exp\left[-\left(\frac{R}{R_{\mathrm{c}}}\right)^{2-\gamma}\right] \tag{4}$$

where $M_{\mathrm{gas}} = 0.09$ solar masses is the total gas mass, $R_{\mathrm{c}} = 160.6$ AU is a characteristic radius, and $\gamma = 1$. The disk temperature structure has a power-law profile with cylindrical radius $R$, and a vertical gradient is imposed at each radius to smoothly connect the cold midplane and warm surface as follows:

$$T(R,Z) = \begin{cases} T_{\mathrm{atm}}(R) + (T_{\mathrm{mid}}(R) - T_{\mathrm{atm}}(R))\cos^2\left(\dfrac{\pi}{2}\dfrac{Z}{Z_{\mathrm{q}}}\right) & \text{if } Z < Z_{\mathrm{q}} \\ T_{\mathrm{atm}}(R) & \text{if } Z \geq Z_{\mathrm{q}} \end{cases} \tag{5}$$

Here, $T_{\mathrm{mid}}(R) = 16.6(R/R_{\mathrm{c}})^{q}$ (in K) and $T_{\mathrm{atm}}(R) = 88.9(R/R_{\mathrm{c}})^{q}$ (in K) where $q = -0.216$ and $Z_{\mathrm{q}} = 97.8(R/R_{\mathrm{c}})^{1.3}$ AU.

Using the parametric disk model described above, we solve the vertical hydrostatic equilibrium equation to obtain an initial density distribution:

$$\rho(R,Z) = \rho(R,0)\frac{c^2(R,0)}{c_{\mathrm{s}}^2(R,Z)}\exp\left[-\int_0^Z \frac{1}{c_{\mathrm{s}}^2(R,Z')}\frac{GM_* Z'}{(R^2+Z'^2)^{3/2}}\mathrm{d}Z'\right] \tag{6}$$

where $\rho$ is the gas density, $c_{\mathrm{s}}$ is the sound speed, $G$ is the gravitational constant, and $M_*$ is the stellar mass. We numerically find the density at each grid cell such that the vertically integrated surface density at the radius becomes the one given in equation (4). With the temperature and density assigned for each grid cell, we compute the initial angular velocity that satisfies the radial force balance taking into account the gas pressure gradient:

$$\Omega = \left[\Omega_K^2 \sin\theta + \frac{1}{\rho r \sin^2\theta}\frac{\partial P}{\partial r}\right]^{1/2} \tag{7}$$

In the above equation $\Omega_{\mathrm{K}} = \sqrt{GM_*/R^3}$ is the Keplerian angular velocity and $P = \rho c_{\mathrm{s}}^2$ is the gas pressure. The initial radial and meridional velocities are set to zero.

The simulation domain extends from $r_{\mathrm{in}} = 32.1$ AU to $r_{\mathrm{out}} = 401.5$ AU in radius and from 0 to 2π in azimuth. In the meridional direction, we include only the upper half of the disk assuming a symmetry across the disk midplane. The meridional domain extends from the midplane ($\theta = \pi/2$) to 20.6° above the midplane ($\theta = \pi/2 - 0.36$). We adopt 256 logarithmically spaced grid cells in the radial direction, 36 uniformly spaced grid cells in the meridional direction, and 635 uniformly spaced grid cells in the azimuthal direction.

We add a $0.5M_{\mathrm{Jup}}$ planet at 87 AU, a $1M_{\mathrm{Jup}}$ planet at 140 AU, and a $2M_{\mathrm{Jup}}$ planet at 237 AU. The planet masses and radial locations are chosen based on previous modelling that successfully reproduced ALMA continuum and CO line observations[6,7,34]. The simulation runs for 1.44 Myr, covering a large fraction of the system's age. This corresponds to 1,000 orbits at $R = R_{\mathrm{c}}$. We linearly increase planet masses during the first 0.144 Myr. We implement a uniform disk viscosity of $\alpha = 10^{-3}$, consistent with the constraints made with molecular line observations of the disk[13,19].

**Additional hydrodynamic simulations.** We carry out the fiducial calculation at a factor of two higher numerical resolution (that is, 512 × 72 × 1,270 grid cells) for a shorter duration of 0.432 Myr. We confirm that the meridional circulation pattern and its strength are not affected by the change in numerical resolution.

While previously inferred disk thermal profile and planet masses successfully reproduce the overall circulatory flow patterns, our fiducial calculation exhibits stronger rotational motions than vertical motion around the gaps compared with those inferred from the observation. We carry out additional calculations with (1) a factor of two smaller planet masses and (2) a factor of two smaller disk surface temperature ($T_{\mathrm{atm}}$ in equation (5)), but these changes result in lowering the level of both rotational and vertical motions. One possible way to have smaller rotational velocity deviations around gaps, while maintaining vertical motions, is if the Reynolds and/or Maxwell stresses in the surface layers are larger than those near the midplane. In such a case, restoring flows near the surface would fill in a gap more efficiently, reducing the pressure gradient across the gap and hence rotation velocity deviations.

The vertical motion from the top of a gap to the planet is, on the other hand, less sensitive to the viscous stress[9]; instead, thermal and magnetic properties within the gap can have a larger influence on the flow pattern and speed[20]. Including ideal and non-ideal magneto-hydrodynamic effects and a more realistic thermal structure via Monte Carlo radiative transfer calculations would be required to examine such a possibility. However, these simulations would require sufficiently high numerical resolution to capture unstable modes of the magneto-rotational instability[35] and long enough integration time for planets to open gaps in a global simulation domain, and are computationally prohibitive for the moment.

## Data availability

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2013.1.00366.S, ADS/JAO.ALMA#2013.1.00601.S and ADS/JAO.ALMA#2016.1.00484.L. The raw data are available from the ALMA archive (http://almascience.nrao.edu/aq/), while the imaged data and scripts are available from the DSHARP website (https://bulk.cv.nrao.edu/almadata/lp/DSHARP/). The Python packages used for the analysis of the data are available via their GitHub repositories: bettermoments (https://github.com/richteague/bettermoments) and eddy (https://github.com/richteague/eddy).

29. Teague, R. eddy. *J. Open Source Softw.* **4**, 1220 (2019).
30. Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G. & Andrae, R. Estimating distance from parallaxes. IV. Distances to 1.33 billion stars in Gaia data release 2. *Astron. J.* **156**, 58 (2018).
31. Teague, R. et al. Measuring turbulence in TW Hydrae with ALMA: methods and limitations. *Astron. Astrophys.* **592**, A49 (2016).
32. Benítez-Llambay, P. & Masset, F. S. FARGO3D: a new GPU-oriented MHD code. *Astrophys. J. Suppl. Ser.* **223**, 11 (2016).
33. Masset, F. FARGO: A fast eulerian transport algorithm for differentially rotating disks. *Astron. Astrophys. Suppl. Ser.* **141**, 165–173 (2000).
34. Liu, S.-F., Jin, S., Li, S., Isella, A. & Li, H. New constraints on turbulence and embedded planet mass in the HD 163296 disk from planet-disk hydrodynamic simulations. *Astrophys. J.* **857**, 87 (2018).
35. Balbus, S. A. & Hawley, J. F. Instability, turbulence, and enhanced transport in accretion disks. *Rev. Mod. Phys.* **70**, 1–53 (1998).

**Acknowledgements** This paper makes use of the following ALMA data: ADS/JAO.ALMA#2013.1.00366.S, ADS/JAO.ALMA#2013.1.00601.S and ADS/JAO.ALMA#2016.1.00484.L. ALMA is a partnership of the European Southern Observatory (ESO; representing its member states), the National Science Foundation (NSF; USA) and the National Institutes of Natural Sciences (Japan), together with the National Research Council (Canada), the National Science Council and the Academia Sinica Institute of Astronomy and Astrophysics (Taiwan), and the Korea Astronomy and Space Science Institute (Korea), in cooperation with Chile. The Joint ALMA Observatory is operated by ESO, Associated Universities, Inc./National Radio Astronomy Observatory (NRAO), and the National Astronomical Observatory of Japan. The NRAO is a facility of the NSF operated under cooperative agreement by Associated Universities, Inc. R.T and E.A.B. acknowledge funding from NSF grant AST-1514670 and NASA grant NNX16AB48G. J.B. acknowledges support from NASA grant NNX17AE31G, and computing resources provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center and by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number ACI-1548562.

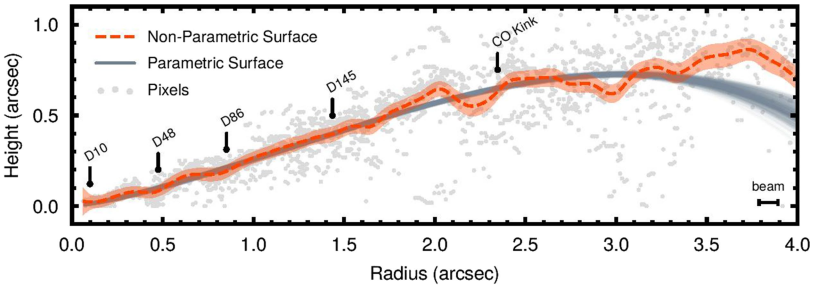
**Author contributions** R.T. devised the method and analysed the data. J.B. ran the hydrodynamic simulations. All authors wrote the manuscript and were participants in the discussion and interpretation of results, determination of the conclusions and revision of the manuscript.

**Competing interests** The authors declare no competing interests.

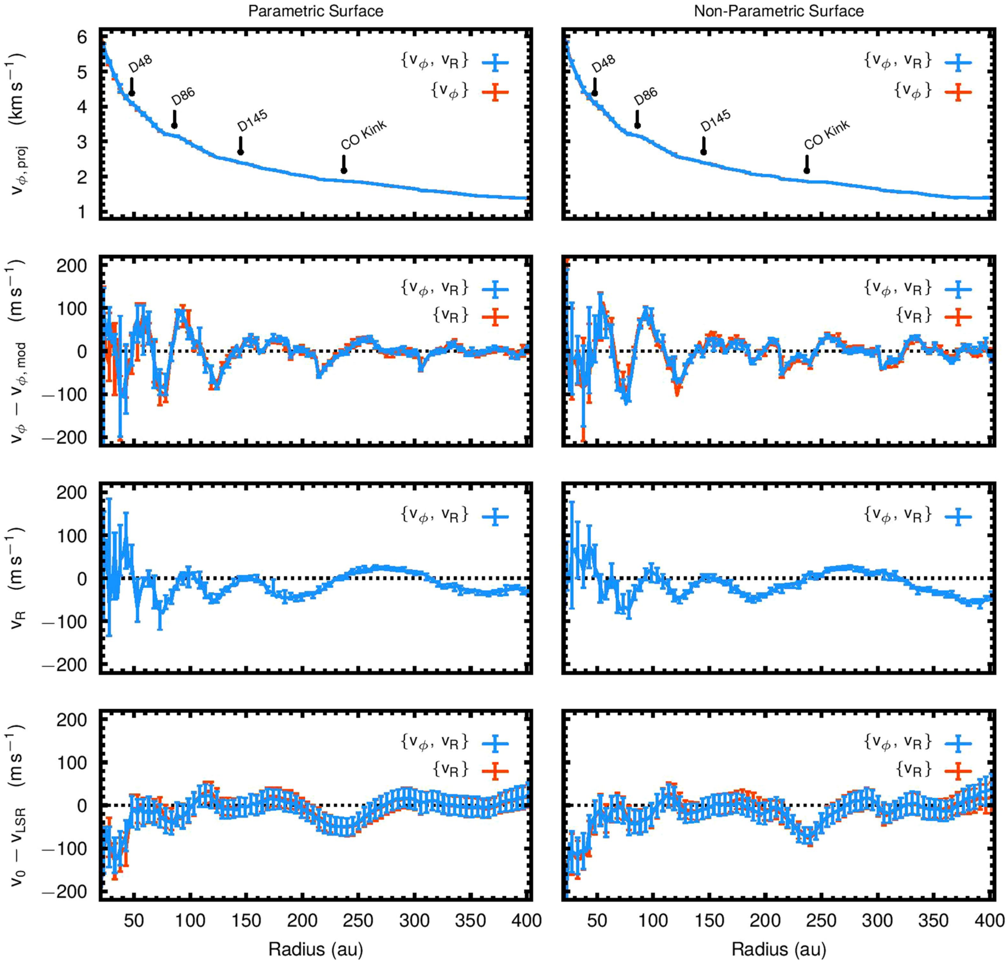
**Additional information**

**Correspondence and requests for materials** should be addressed to R.T.

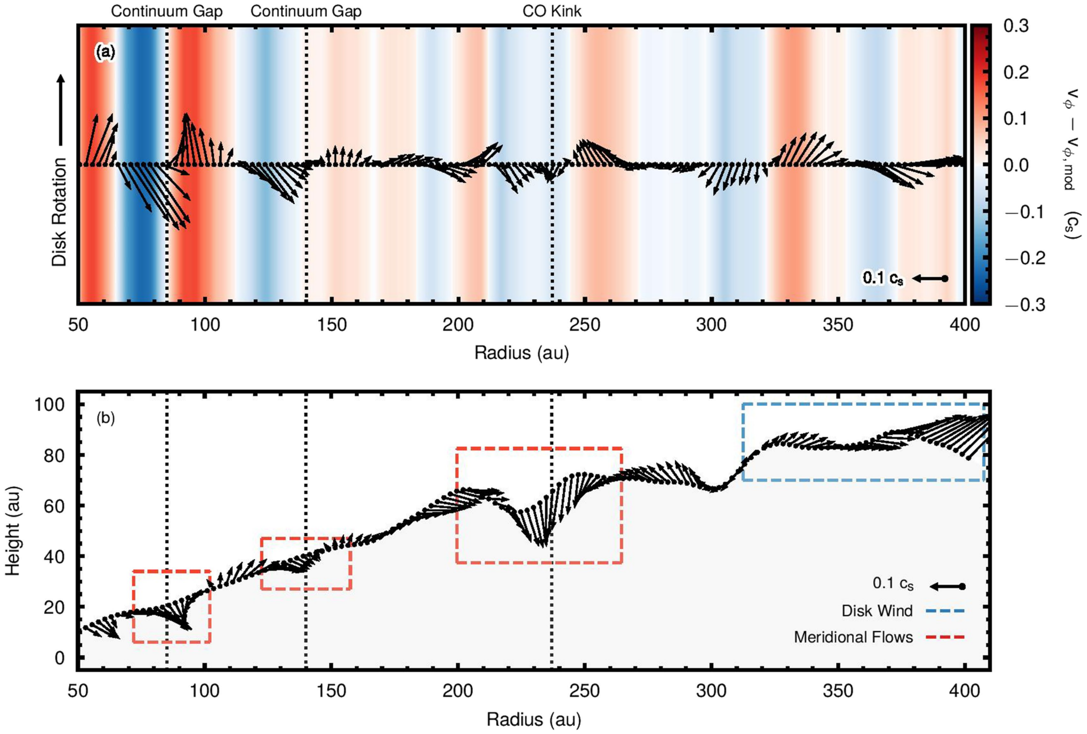
**Reprints and permissions information** is available at http://www.nature.com/reprints.



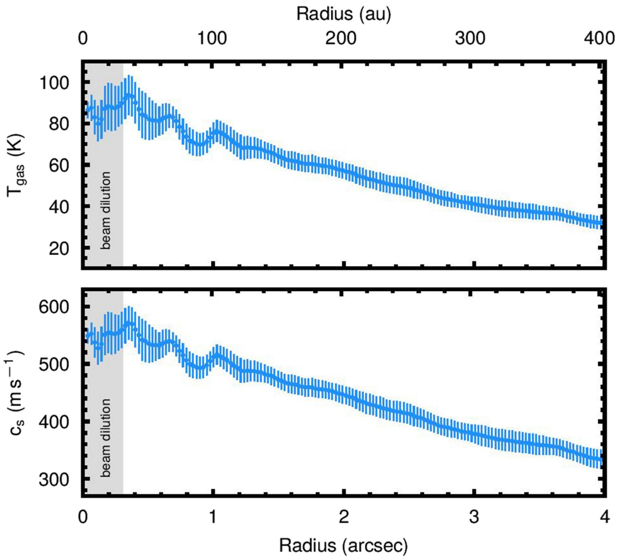
**Extended Data Fig. 1 | A comparison of emission heights inferred for the $^{12}$CO emission.** The grey points in the background represent individual measurements following ref. [15], while the red contour shows the Gaussian process model of this surface including 1$\sigma$ uncertainties, as described in ref. [6]. Grey lines are random samples from the parametric fit derived from modelling the line-of-sight velocity map, with their spread demonstrating the 1$\sigma$ uncertainties. The dust gap locations[5,28] (D10 to D145) and radial location of the velocity perturbation found[7] in $^{12}$CO, the 'CO kink', are marked. The major axis of the beam is shown for scale at bottom right.



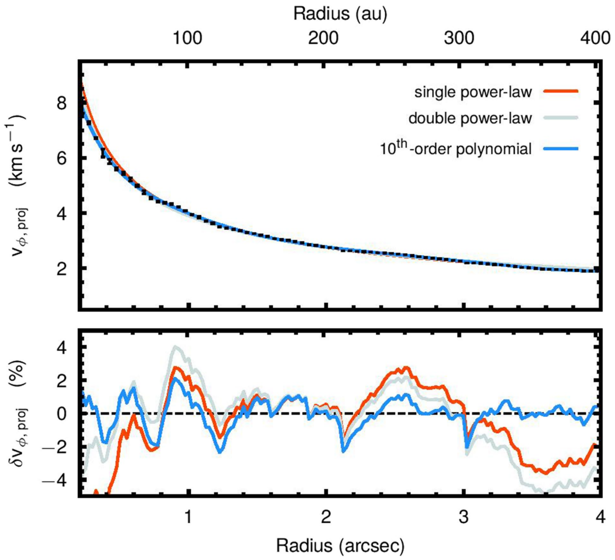
**Extended Data Fig. 2 | Measured velocity structure of HD 163296.** Top row, projected rotation velocity $v_{\phi,\mathrm{proj}}$; second row, the residual from the 10th-order polynomial fit to $v_\phi$ to highlight the small-scale structure; third row, the $v_R$ values; and fourth row, the deviation in the shifted and aligned line centre from the systemic velocity. Left and right columns show results using the parametric and non-parametric emission surface, respectively. Velocities in the lower three rows have been corrected for projection effects assuming $i = 47.6°$. Blue error bars show the inferred velocities assuming both $v_R$ and $v_\phi$ components, while red error bars assume $v_R = 0$ m s$^{-1}$.



**Extended Data Fig. 3 | Measured velocity structure of the gas in the disk around HD 163296. a**, **b**, As Fig. 2 but using the non-parametric emission surface to deproject the data. Structure in the emission height outside 3″ is due to higher noise in the data, as described in the text.



**Extended Data Fig. 4 | Gas temperature and sound speed.** The figure shows the derived gas temperature ($T_{gas}$, top panel) and the derived gas sound speed ($c_s$, bottom panel) as a function of radius. Error bars show the $1\sigma$ uncertainty. The drop in these values in the inner ~30 AU (shaded area) is due to beam dilution.



**Extended Data Fig. 5 | Impact of the choice of velocity baseline.** The figure shows how the choice of $v_{\phi,\mathrm{mod}}$ affects the residuals from $v_\phi$, as in the second row of Extended Data Fig. 2. The top panel shows the different underlying models compared to the unprojected data, while the bottom panel shows the residual between the model and the observations. Regardless of the $v_{\phi,\mathrm{mod}}$ chosen, the structure in $\delta v_\phi = (v_\phi - v_{\phi,\mathrm{mod}})/v_\phi$ persists.

LETTER

https://doi.org/10.1038/s41586-019-1568-6

# Supersolid symmetry breaking from compressional oscillations in a dipolar quantum gas

L. Tanzi[1,2,3,6], S. M. Roccuzzo[4,5,6], E. Lucioni[1,2,3], F. Famà[1], A. Fioretti[1], C. Gabbanini[1], G. Modugno[1,2,3]*, A. Recati[4,5]* & S. Stringari[4,5]

**Supersolids are exotic materials combining the frictionless flow of a superfluid with the crystal-like periodic density modulation of a solid. The supersolid phase of matter was predicted 50 years ago[1–3] for solid helium[4–8]. Ultracold quantum gases have recently been made to exhibit periodic order typical of a crystal, owing to various types of controllable interaction[9–13]. A crucial feature of a *D*-dimensional supersolid is the occurrence of $D+1$ gapless excitations, reflecting the Goldstone modes associated with the spontaneous breaking of two continuous symmetries: the breaking of phase invariance, corresponding to the locking of the phase of the atomic wave functions at the origin of superfluid phenomena, and the breaking of translational invariance due to the lattice structure of the system. Such modes have been the object of intense theoretical investigations[1,14–18], but they have not yet been observed experimentally. Here we demonstrate supersolid symmetry breaking through the appearance of two distinct compressional oscillation modes in a harmonically trapped dipolar Bose–Einstein condensate, reflecting the gapless Goldstone excitations of the homogeneous system. We observe that the higher-frequency mode is associated with an oscillation of the periodicity of the emergent lattice and the lower-frequency mode characterizes the superfluid oscillations. This work also suggests the presence of two separate quantum phase transitions between the superfluid, supersolid and solid-like configurations.**

Despite decades of investigation, supersolidity in solid helium has not yet been demonstrated[4–8]. Quantum gases with spin–orbit coupling[9,19], cavity-mediated interactions[10,20] and long-range dipolar interactions[11–13,17,18,21–24] are emerging as interesting alternatives. A major difference with respect to solid helium is that the lattice structure of these configurations is built by clusters rather than by single atoms, thereby naturally allowing for the emergence of coherence effects typical of a superfluid. The recently observed supersolid stripe phase in elongated dipolar Bose–Einstein condensates[11–13] is particularly appealing, since the lattice-like structure is determined directly by the atom–atom interactions and not by light-mediated interactions, allowing the lattice to be deformable. However, how to reveal spontaneous symmetry breaking in such trapped, inhomogeneous systems and how to study their lattice dynamics are still open questions.

In this work, we demonstrate, both theoretically and experimentally, that the peculiar symmetry breaking of the dipolar supersolid can be revealed by studying the collective oscillations of the system in the trap. Low-frequency compressional modes emerge naturally from the hydrodynamic equations of superfluids and have been studied extensively in the context of trapped Bose–Einstein condensates[25–27]. The hydrodynamic equations are the direct consequence of the locking of the phase of the order parameter and are hence peculiar to superfluid systems at low temperature[28]. We now discover that, when the system is driven from the usual superfluid to the supersolid regime by tuning the interactions, the lowest compressional mode—the so-called axial breathing mode—bifurcates into two distinct excitations, similar to the bifurcation of the gapless Goldstone excitations expected for a homogeneous supersolid. By further varying the interactions, one of the two modes disappears, and the system shows a second transition from the supersolid to a droplet crystal regime.

The system we study is initially an ordinary Bose–Einstein condensate (BEC) of strongly magnetic dysprosium (Dy) atoms (see Fig. 1a), held in a harmonic potential with frequencies $\omega_x$, $\omega_y$ and $\omega_z$, with the dipoles aligned along *z* by a magnetic field *B*; the potential is elongated in the *x* direction. At a mean-field level, the atoms interact with a short-range repulsion parametrized by the s-wave scattering length $a_s$, and with a long-range dipolar interaction parametrized by the dipolar length $a_{dd}$, which is repulsive along the *x* and *y* directions and attractive along the *z* direction. By changing $a_s$ via a Feshbach resonance, we can control the relative interaction strength, $\epsilon_{dd} = a_{dd}/a_s$. By increasing $\epsilon_{dd}$ from the BEC side, a roton minimum develops in the excitation spectrum[29], the energy gap of which becomes softer and softer. Eventually, at the critical value $\epsilon_{dd} = 1.38$, an instability develops and a periodic density modulation appears along the *x* direction, with a characteristic momentum set by the trap confinement along *z*, $k_r \approx (\hbar/m\omega_z)^{-1/2}$ (ref. [30]). The modulation would not be stable at the mean-field level but is stabilized by the Lee–Huang–Yang interaction, which is the leading term accounting for beyond-mean-field quantum fluctuations[31]. For a narrow range of values of $\epsilon_{dd}$, the system shows both density modulation and phase coherence, which are the prerequisites for supersolidity[11–13]. Owing to the presence of three-body recombination, enhanced by the emergence of high-density regions, the lifetime of the supersolid is limited to hundreds of milliseconds. For larger values of $\epsilon_{dd}$, the system enters an incoherent droplet crystal regime[32,33].

Theoretically, we study the system by means of a density functional theory with a local density approximation for the equation of state, which includes the Lee–Huang–Yang interaction correction. We consider the ideal scenario of zero temperature and the absence of three-body recombination. We prepare the system in its ground state at a specific value of $\epsilon_{dd}$. We induce an axial breathing mode along the *x* direction, through a sudden modification of the harmonic potential, and we solve the corresponding time-dependent equation, which takes the form of a generalized time-dependent Gross–Pitaevskii equation. We then monitor the evolution of the second moment $x^2$ of the in-trap density distribution, which, as shown in Fig. 1b, presents a perfectly sinusoidal oscillation. In the absence of the dipolar interaction, the mean-field solution for the breathing mode frequency can be analytically obtained by solving the hydrodynamic equations of superfluids[25], yielding $\omega = \sqrt{5/2}\,\omega_x$, in excellent agreement with experimental observations[27]. In the presence of dipolar interaction, we find a breathing frequency in agreement with hydrodynamic calculations[34], a value that we show extends beyond the mean-field region of stability of the BEC. The same value is obtained by a sum rule calculation, which provides a rigorous upper bound for the mode frequency (see Methods), as well as the solutions of the Bogoliubov equations[35,36].

---

[1]CNR-INO, Sede Secondaria di Pisa, Pisa, Italy. [2]LENS, Università di Firenze, Florence, Italy. [3]Dipartimento di Fisica e Astronomia, Università di Firenze, Florence, Italy. [4]CNR-INO, BEC Center, Trento, Italy. [5]Dipartimento di Fisica, Università di Trento, Trento, Italy. [6]These authors contributed equally: L. Tanzi, S. M. Roccuzzo. *e-mail: modugno@lens.unifi.it; alessio.recati@unitn.it

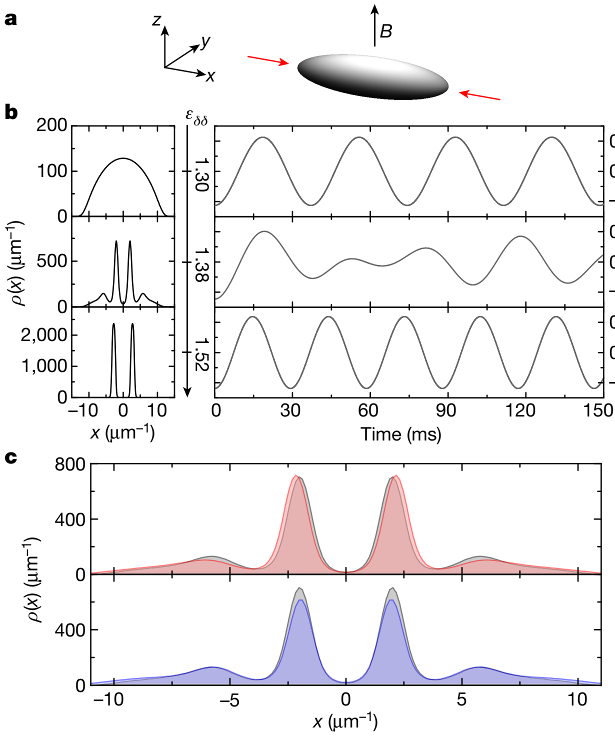


**Fig. 1 | Doubling of the axial breathing mode in the supersolid regime (theory). a**, Geometry of the system. The axial breathing of the BEC leads mainly to a variation of the width along the *x* direction (arrows). **b**, Theoretical calculations of the stationary in-trap density distributions along the *x* direction (left) and of the time-dependent oscillations in its normalized width, $\sigma_x(t)=\sqrt{x(t)^2/x(0)^2}$. Three representative cases are shown: a standard BEC (top row, $\epsilon_{\rm dd}=1.30$), the supersolid regime (middle row, $\epsilon_{\rm dd}=1.38$) and the droplet crystal regime (bottom row, $\epsilon_{\rm dd}=1.52$). Whereas the BEC and the crystal feature only one oscillation frequency, the supersolid regime clearly shows the beating of two frequencies, proving that the periodic modulation of the density is accompanied by the appearance of a new compressional mode. **c**, Different frames of the time evolution ($t=0$, grey; $t=20$ ms, blue; $t=50$ ms, red) representative of the character of the two modes appearing in the beatings in the middle panels of **b**: the upper panel (lower panel) shows the lattice deformation (amplitude modulation) associated with the higher (lower) frequency.

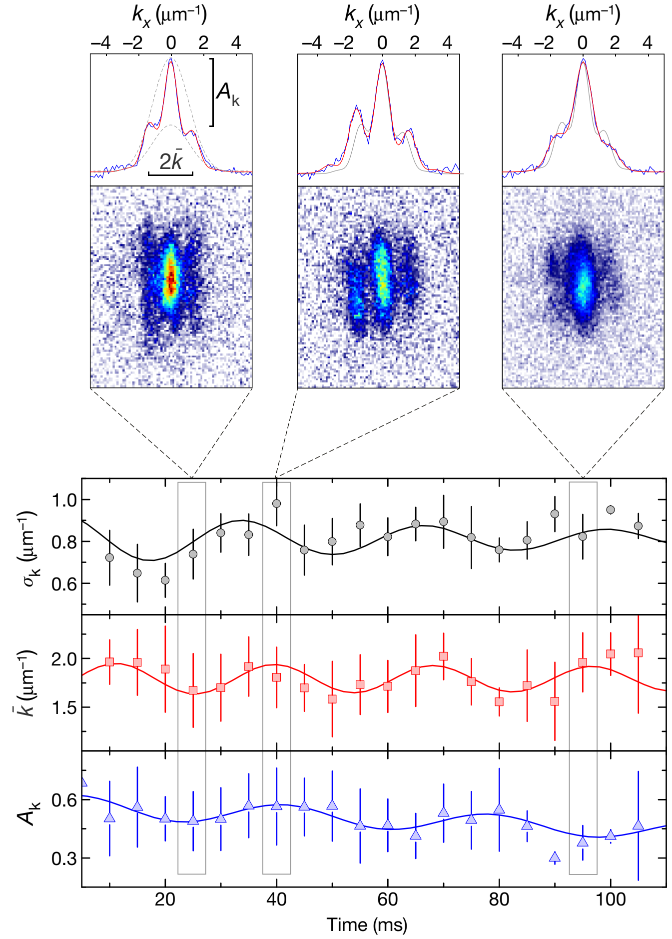


**Fig. 2 | Doubling of the axial breathing mode in the supersolid regime (experiment).** Experimental oscillation modes in the supersolid regime ($\epsilon_{\rm dd}=1.38$), by monitoring the width of the distributions in momentum space, $\sigma_k(t)=\sqrt{k_x(t)^2}$ (top panel, grey), the spacing of the side peaks $\bar{k}(t)$ (middle panel, red) and the modulation amplitude $A_k(t)$ (bottom panel, blue). Experimental data (circles, squares and triangles) are fitted with damped sinusoids (lines), to measure the dominant frequencies, which are $\omega/\omega_x=1.45(8)$, 1.66(10) and 1.27(12) for $\sigma_k$, $\bar{k}$ and $A_k$, respectively. The insets show samples of the experimental false-colour distributions in the ($k_x$, $k_y$) plane and the related fits of $n(k_x)$ with a two-slit model to measure $\bar{k}$ and $A_k$; grey lines show the fit to the first distribution, for comparison. Error bars represent the standard deviation of 4–8 measurements.

For values of $\epsilon_{\rm dd}\geq 1.38$, the equilibrium profiles exhibit a lattice structure, a signal of supersolidity[11] (see Fig. 1b). Remarkably, the corresponding time-dependent solutions reveal a beating that reflects the presence of two oscillation modes. The nature of the two modes is identified through an analysis of the time evolution of the lattice period, which is dominated by the higher-frequency mode, as well as an analysis of the lattice amplitude, which is dominated by the lower-frequency mode (see Methods). Figure 1c shows two examples of axial density distributions obtained at different evolution times, showing the dominant character of the two modes. We note that the bimodal oscillation emerges clearly only if we consider small-amplitude oscillations (<5%) and tends to disappear for larger amplitudes, confirming that the two modes are strongly coupled, as is easily seen from the hydrodynamic-like equations for the untrapped uniform gas[14]. When $\epsilon_{\rm dd}$ is further increased into the droplet regime, the lower mode disappears and only the upper one remains.

In the experiment, differing from the theoretical approach, the system is prepared in the BEC regime and then slowly brought to the supersolid regime by increasing $\epsilon_{\rm dd}$. We reveal the oscillations by monitoring the momentum distribution $n(k_x)$ along $x$, after a free expansion (see Methods). Simulating the initial expansion is challenging, so we cannot exactly relate the experimental and theoretical observables; however, the oscillation frequencies are not affected by the expansion. In the BEC regime, we can excite the axial breathing oscillation with a controlled amplitude by suddenly changing $a_{\rm s}$. We monitor the second moment of the distribution, $k_x^{\,2}$, the oscillation of which features a weak damping caused by finite-temperature effects. The same finite-temperature analysis confirms that the oscillation frequency of the BEC is distinct from that of a classical gas ($\omega=2\,\omega_x$) and is therefore a consequence of the breaking of phase invariance (see Methods).

The transition to the supersolid regime is signalled by the appearance of side peaks in $n(k_x)$, which reflect the in-trap lattice[11] (see Fig. 2). We observe that the instability itself naturally triggers the axial oscillation with a typical amplitude of 10%, which therefore represents the minimum amplitude in that regime. Owing to the decay of the density modulation via three-body losses, the observation time is limited to a few periods, reducing the precision of the measurements. Differing from the theoretical prediction, in the time evolution of $k_x^{\,2}$ we observe only a single mode, with a lower frequency than the BEC one. We attribute the absence of two beating frequencies to the relatively large amplitude of the oscillations in the experiment, which results in a mixing of the two modes, owing to their coupled nature. We can, however,

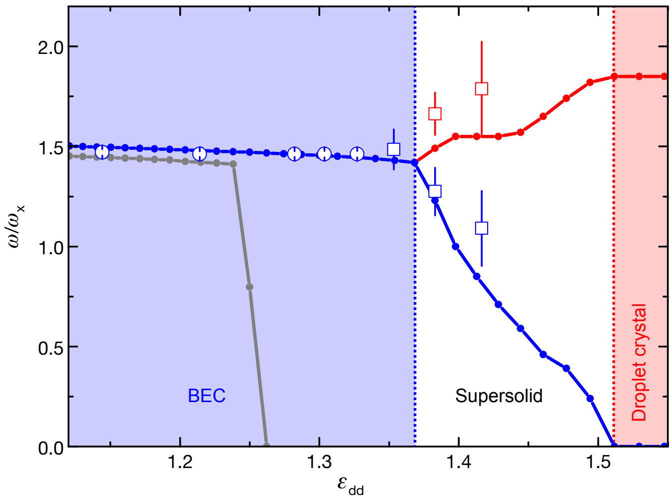


**Fig. 3 | Axial mode frequencies from BEC to supersolid and droplet crystal.** The shaded regions identify the different phases, separated by dotted lines determined as described in the text. The dipolar character of the system is varied through the parameter $\epsilon_{\rm dd} = a_{\rm dd}/a_{\rm s}$. Dotted lines are the theoretically predicted frequencies including the Lee–Huang–Yang interaction energy term that is due to quantum fluctuations (blue and red) or excluding it (grey). Large circles and squares are experimentally measured frequencies: in the BEC regime (circles) the oscillation is induced by quenching $\epsilon_{\rm dd}$; in the supersolid regime (squares) the instability naturally triggers the oscillation. Colours indicate the dominant character of the two modes in the supersolid: superfluid-related (blue) or lattice-related (red). Error bars are one standard deviation. In the experiment, $\epsilon_{\rm dd}$ has a calibration uncertainty of 3%.

monitor two different observables, more directly related to the two modes found in the theory: the spacing $\bar{k}$ of the side peaks, which is the inverse lattice period, and the modulation amplitude $A_k$, which is associated with the depth of the lattice. As shown in Fig. 2, the two observables feature clear sinusoidal oscillations, with different frequencies: $\bar{k}$ oscillates with a higher frequency and $A_k$ oscillates with a lower frequency, in agreement with the theoretical predictions.

In Fig. 3 both the predicted and measured frequencies versus $\epsilon_{\rm dd}$ are reported, together with the phase diagram. The appearance of a density modulation marks the transition to the supersolid regime (vertical blue dotted line). The transition from the supersolid to the droplet crystal regime (vertical red dotted line) is identified with the zero crossing of the calculated chemical potential after subtracting the contribution of the external harmonic potential, which provides the onset of the formation of self-bound droplets. In the supersolid, the higher-frequency mode is clearly related to the lattice deformations and its frequency increases owing to the dipolar repulsion between neighbouring density maxima. The lower mode is instead related to the compressional oscillation of the superfluid component; its downward frequency shift can be justified as an effective mass acquired by the atoms moving through the lattice, corresponding to a reduction of the superfluid fraction. The lower mode eventually disappears as the system enters the droplet crystal regime, in analogy with the behaviour of the corresponding Goldstone mode in uniform systems[1,14,17,22]. The upper mode instead approaches the value characteristic of a solid phase of incoherent droplets (see Methods). It is worth noticing that the changes in the dynamics coincide with the predicted transition lines.

The agreement of experiment with theory on the BEC side is remarkable. The comparison with theoretical results without the Lee–Huang–Yang interaction term shows a clear stabilizing effect of quantum fluctuations already on the BEC side. In the supersolid regime, the splitting of the experimental oscillation frequencies agrees qualitatively with the theory. For larger $\epsilon_{\rm dd}$, even before the theoretical prediction for the transition to the droplet crystal, the $\bar{k}$ and $A_k$ modes are no longer visible in the experiment because $n(k)$ becomes incoherent (see Methods). The spontaneous excitation of the axial mode observed in the experiment at the BEC–supersolid transition reveals a release of excess energy, suggesting a first-order phase transition, in agreement with theoretical predictions[17–19,22,24].

In conclusion, the bifurcation of the lowest compressional mode of a harmonically trapped dipolar supersolid provides evidence of the simultaneous breaking of two continuous symmetries, in analogy with the gapless Goldstone modes predicted for a homogeneous supersolid. The nature of the mode associated with the broken translational symmetry demonstrates the compressibility of the crystal structure of the dipolar supersolid, in analogy with the hypothesized helium supersolid[1] and in contrast with the incompressible cavity supersolid[10,20,37]. This access to the compressional modes opens up exploration of the non-trivial competition between the superfluid and crystalline features exhibited in this type of supersolid, in which two separate quantum phase transitions can be accessed by tuning a single interaction parameter, $\epsilon_{\rm dd}$. For example, a detailed analysis of the mode splitting at the BEC–supersolid transition might confirm that the transition is of first order, in analogy with the discontinuity in the phonon velocities of the Goldstone modes[22]. Working with systems having reduced losses[13] and larger sizes, one might also explore the transition between the supersolid and the droplet crystal, at which the theory predicts the disappearance of the lower-frequency mode. Although the existence of two Goldstone modes is a manifestation of superfluidity, our work does not exhaust the assessment of superfluidity in supersolid dipolar gases. The study of other important consequences of superfluidity—such as the occurrence of permanent currents and of a critical velocity, the reduction of the moment of inertia, as well as the existence of quantized vortices—requires the implementation of different geometrical configurations and different experimental techniques that will be developed in future work.

We note that after completing the present study, we became aware of related investigations by the Stuttgart and Innsbruck groups[38,39].

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1568-6.

Received: 6 June 2019; Accepted: 29 August 2019;
Published online 9 September 2019.

1. Andreev, A. F. & Lifshitz, I. M. Quantum theory of defects in crystals. *Sov. Phys. JETP* **29**, 1107–1113 (1969).
2. Chester, G. V. Speculations on Bose–Einstein condensation and quantum crystals. *Phys. Rev. A* **2**, 256–258 (1970).
3. Leggett, A. J. Can a solid be "superfluid"? *Phys. Rev. Lett.* **25**, 1543–1546 (1970).
4. Boninsegni, M. & Prokof'ev, N. V. *Colloquium*: Supersolids: what and where are they? *Rev. Mod. Phys.* **84**, 759 (2012).
5. Kim, E. & Chan, M. H. W. Probable observation of a supersolid helium phase. *Nature* **427**, 225–227 (2004).
6. Balibar, S. The enigma of supersolidity. *Nature* **464**, 176–182 (2010).
7. Kim, D. Y. & Chan, M. H. W. Absence of supersolidity in solid helium in porous Vycor glass. *Phys. Rev. Lett.* **109**, 155301 (2012).
8. Nyéki, J. et al. Intertwined superfluid and density wave order in two-dimensional $^4$He. *Nat. Phys.* **13**, 455–459 (2017).
9. Li, J.-R. et al. A stripe phase with supersolid properties in spin–orbit-coupled Bose–Einstein condensates. *Nature* **543**, 91–94 (2017).
10. Léonard, J., Morales, A., Zupancic, P., Esslinger, T. & Donner, T. Supersolid formation in a quantum gas breaking a continuous translational symmetry. *Nature* **543**, 87–90 (2017).
11. Tanzi, L. et al. Observation of a dipolar quantum gas with metastable supersolid properties. *Phys. Rev. Lett.* **122**, 130405 (2019).
12. Böttcher, F. et al. Transient supersolid properties in an array of dipolar quantum droplets. *Phys. Rev. X* **9**, 011051 (2019).
13. Chomaz, L. et al. Long-lived and transient supersolid behaviors in dipolar quantum gases. *Phys. Rev. X* **9**, 021012 (2019).
14. Josserand, C., Pomeau, Y. & Rica, S. Coexistence of ordinary elasticity and superfluidity in a model of a defect-free supersolid. *Phys. Rev. Lett.* **98**, 195301 (2007).
15. Watanabe, H. & Brauner, T. Spontaneous breaking of continuous translational invariance. *Phys. Rev. D* **85**, 085010 (2012).

16. Kunimi, M. & Kato, Y. Mean-field and stability analyses of two-dimensional flowing soft-core bosons modeling a supersolid. *Phys. Rev. B* **86**, 060510(R) (2012).
17. Saccani, S., Moroni, S. & Boninsegni, M. Excitation spectrum of a supersolid. *Phys. Rev. Lett.* **108**, 175301 (2012).
18. Roccuzzo S. M. & Ancilotto, F. Supersolid behavior of a dipolar Bose-Einstein condensate confined in a tube. *Phys. Rev. A* **99**, 041601(R) (2019).
19. Li, Y., Pitaevskii, L. P. & Stringari, S. Quantum tricriticality and phase transitions in spin-orbit-coupled Bose–Einstein condensates. *Phys. Rev. Lett.* **108**, 225301 (2012).
20. Lang, J., Piazza, F. & Zwerger, W. Collective excitations and supersolid behavior of bosonic atoms inside two crossed optical cavities. *New J. Phys.* **19**, 123027 (2017).
21. Pupillo, G., Micheli, A., Boninsegni, M., Lesanovsky, I. & Zoller, P. Strongly correlated gases of Rydberg-dressed atoms: quantum and classical dynamics. *Phys. Rev. Lett.* **104**, 223002 (2010).
22. Macrì, T., Maucher, F., Cinti, F., & Pohl, T. Elementary excitations of ultracold soft-core bosons across the superfluid-supersolid phase transition. *Phys. Rev. A* **87**, 061602 (2013).
23. Ancilotto, F., Rossi, M. & Toigo, F. Supersolid structure and excitation spectrum of soft-core bosons in three dimensions. *Phys. Rev. A* **88**, 033618 (2013).
24. Lu, Z.-K., Li, Y., Petrov, D. S. & Shlyapnikov, G. V. Stable dilute supersolid of two-dimensional dipolar bosons. *Phys. Rev. Lett.* **115**, 075303 (2015).
25. Stringari, S. Collective excitations of a trapped Bose-condensed gas. *Phys. Rev. Lett.* **77**, 2360–2363 (1996).
26. Jin, D. S., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Collective excitations of a Bose-Einstein condensate in a dilute gas. *Phys. Rev. Lett.* **77**, 420–423 (1996).
27. Mewes, M. O. et al. Collective excitations of a Bose-Einstein condensate in a magnetic trap. *Phys. Rev. Lett.* **77**, 988–991 (1996).
28. Pitaevskii, L. & Stringari, S. *Bose-Einstein Condensation and Superfluidity* (Oxford Univ. Press, 2011).
29. Petter, D. et al. Probing the roton excitation spectrum of a stable dipolar Bose gas. *Phys. Rev. Lett.* **122**, 183401 (2019).
30. Chomaz, L. et al. Observation of roton mode population in a dipolar quantum gas. *Nat. Phys.* **14**, 442–446 (2018).
31. Lima, A. R. P. & Pelster, A. Beyond mean-field low-lying excitations of dipolar Bose gases. *Phys. Rev. A* **86**, 063609 (2012).
32. Kadau, H. et al. Observing the Rosensweig instability of a quantum ferrofluid. *Nature* **530**, 194–197 (2016).
33. Wenzel, M., Böttcher, F., Langen, T., Ferrier-Barbut, I. & Pfau, T. Striped states in a many-body system of tilted dipoles. *Phys. Rev. A* **96**, 053630 (2017).
34. van Bijnen, R. M. W., Parker, N. G., Kokkelmans, S. J. J. M. F., Martin, A. M. & O'Dell, D. H. J. Collective excitation frequencies and stationary states of trapped dipolar Bose-Einstein condensates in the Thomas-Fermi regime. *Phys. Rev. A* **82**, 033612 (2010).
35. Ronen, S., Bortolotti, D. C. E. & Bohn, J. L. Bogoliubov modes of a dipolar condensate in a cylindrical trap. *Phys. Rev. A* **74**, 013623 (2006).
36. Blakie, P.E., Baillie, D. & Bisset, R. N. Roton spectroscopy in a harmonically trapped dipolar Bose-Einstein condensate. *Phys. Rev. A* **86**, 021604 (2012).
37. Léonard, J., Morales, A., Zupancic, P., Donner, T. & Esslinger, T. Monitoring and manipulating Higgs and Goldstone modes in a supersolid quantum gas. *Science* **358**, 1415–1418 (2017).
38. Guo, M. et al. The low-energy Goldstone mode in a trapped dipolar supersolid. *Nature* https://doi.org/10.1038/s41586-019-1569-5 (2019).
39. Natale, G. et al. Excitation spectrum of a trapped dipolar supersolid and its experimental evidence. *Phys. Rev. Lett.* **123**, 050402 (2019).

## METHODS

**Real-time simulations.** Our theoretical predictions are based on the numerical solutions of a time-dependent density functional equation, which takes the form of an extended Gross–Pitaevskii equation for the condensate wavefunction $\phi(\mathbf{r}, t)$:

$$i\hbar\frac{\partial\phi(\mathbf{r},t)}{\partial t}=\left[\frac{-\hbar^2}{2m}\nabla^2+V_t(\mathbf{r})+\int \mathrm{d}\mathbf{r}'(g\delta(\mathbf{r}-\mathbf{r}')+V_{\mathrm{dd}}(\mathbf{r}-\mathbf{r}'))|\phi(\mathbf{r}',t)|^2\right.$$
$$\left.+\gamma(\epsilon_{\mathrm{dd}})\ |\phi(\mathbf{r},t)|^3\right]\phi(\mathbf{r},t)$$

Here, $V_{\mathrm{t}}(\mathbf{r})=\frac{1}{2}m(\omega_x^2x^2+\omega_y^2y^2+\omega_z^2z^2)$ is the harmonic trapping potential and $m$ is the atomic mass. The contact interaction is expressed by $g\delta(\mathbf{r}-\mathbf{r}')$, with $g=\frac{4\pi\hbar^2a_{\mathrm{s}}}{m}$, where $a_{\mathrm{s}}$ is the s-wave scattering length. $V_{\mathrm{dd}}(\mathbf{r})=\frac{\mu_0\mu^2}{4\pi}\frac{1-3\cos^2\theta}{|\mathbf{r}|^3}$ is the dipole–dipole interaction between two magnetic dipoles aligned along the $z$ axis, with $\mu$ the atomic dipole moment, $\mu_0$ the vacuum permeability and $\theta$ the angle between the vector $\mathbf{r}$ and $z$. The dipolar parameter $\epsilon_{\mathrm{dd}}=a_{\mathrm{dd}}/a_{\mathrm{s}}$ (where $a_{\mathrm{dd}}=\frac{\mu_0\mu^2m}{12\pi\hbar^2}$ is the dipolar length) gives the relative strength of the (partially) attractive dipole–dipole interaction, and the repulsive contact interaction. The last term, $\gamma(\epsilon_{\mathrm{dd}})=\frac{32}{3\sqrt{\pi}}ga_{\mathrm{s}}^{3/2}F(\epsilon_{\mathrm{dd}})$, is the beyond-mean-field Lee–Huang–Yang interaction correction, with $F(\epsilon_{\mathrm{dd}})=\frac{1}{2}\int \mathrm{d}\theta\ \sin\theta[1+\epsilon_{\mathrm{dd}}(3\cos^3\theta-1)]^{5/2}$ (ref. [31]). In the present density functional approach, the local atom density is identified as $n(\mathbf{r},t)=|\phi(\mathbf{r},t)|^2$.

The equilibrium density configuration is found by fixing the equation parameters to the values available in the experiments, and then evolving the extended Gross–Pitaevskii equation in imaginary time. We consider $N=35{,}000$ atoms of $^{162}$Dy, for which $a_{\mathrm{dd}}=130a_0$ (where $a_0$ is the Bohr radius). The trapping frequencies are $\omega_{x,y,z}=2\pi(18.5,\ 53,\ 81)$ Hz, and $a_{\mathrm{s}}$ is changed to tune the value of $\epsilon_{\mathrm{dd}}$.

The axial breathing mode is excited by finding the equilibrium configuration in the presence of a perturbation of the form $\hat{H}=-\lambda\hat{x}^2$, where $\lambda$ is a small parameter, and then evolving the extended Gross–Pitaevskii equation in real time with $\lambda=0$. During the evolution, we keep track of the time dependence of the second moment of the in situ density distribution, defined as $\sigma_x(t)=\sqrt{x^2(t)/x^2(0)}$, with $x^2(t)=\int \mathrm{d}\mathbf{r}x^2n(\mathbf{r},t)$. This quantity shows a simple harmonic oscillation in the BEC and in the droplet crystal phase, whereas a beating between two frequencies clearly emerges in the supersolid regime.

**Theoretical mode assignment in the supersolid phase.** In the supersolid phase, we also record the time evolution of the relative distance $d(t)$ between the peaks of the in situ density distribution, as well as their amplitude $A(t)$. The results are shown in Extended Data Fig. 1, together with their Fourier transforms $\widetilde{d}(\omega)$ and $\widetilde{A}(\omega)$, respectively. The dominant peak in $\widetilde{d}(\omega)$ corresponds to the higher of the two frequencies that contribute to the beating observed in $\sigma_x(t)$, while the dominant peak in $\widetilde{A}(\omega)$ corresponds to the lower one. This allows us to identify the higher-frequency mode as the one associated with a lattice deformation, and the lower-frequency mode as the one associated with the compression of the superfluid.

**Sum rules.** A further insight into the excitation spectrum of the system is provided by the calculation of the sum rule $m_1/m_{-1}$ (ref. [26]), where $m_1$ and $m_{-1}$ are, respectively, the energy-weighted and inverse-energy-weighted moments of the dynamic structure factor, relative to the operator $\hat{x}^2$. Explicit calculation of the two sum rules[40] gives the rigorous upper bound $\omega^2=-2\frac{x^2}{\mathrm{d}x^2/\mathrm{d}\omega_x^2}$ for the square of the frequency of the excited mode, which can be extracted from static calculations. In the BEC regime, we find that the upper bound matches with high precision the frequency of the axial breathing mode calculated by solving the Bogoliubov–de Gennes equations and by real-time simulations, so that the linear response of the system to the operator $\hat{x}^2$ is exhausted only by this mode. In the supersolid regime, instead, the sum rule result lies between the two beating frequencies, so that in this phase the operator excites the higher-energy modes as well.

**Classical model for droplet oscillation.** For the value $\epsilon_{\mathrm{dd}}=1.53$, the equilibrium configuration of the system is given by two self-bound droplets, whose relative distance is fixed by the external trap. We find that the axial mode excited by $\hat{x}^2$ is just an out-of-phase harmonic oscillation of the two droplets without any substantial deformation of their density profiles. To get a better physical insight into the nature of the oscillations in this regime, we study their frequency by treating the two droplets as classical distributions of dipoles. Neglecting their internal structure, we write the energy functional in the form $E[n_1,n_2]=E_{\mathrm{dd}}[n_1,n_2]+E_{\mathrm{trap}}[n_1]+E_{\mathrm{trap}}[n_2]$, where $n_1(\mathbf{r})$ and $n_2(\mathbf{r})$ are the density distributions of the two droplets calculated by solving the extended Gross–Pitaevskii equation, while $E_{\mathrm{dd}}[n_1,n_2]=\frac{1}{2}\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2V_{\mathrm{dd}}(\mathbf{r}_1-\mathbf{r}_2)n_1(\mathbf{r}_1)n_2(\mathbf{r}_2)$ is their dipole–dipole interaction energy and $E_{\mathrm{trap}}n_i=\int \mathrm{d}\mathbf{r}V_t(\mathbf{r})n_i(\mathbf{r})$ is the energy term due to the trapping potential. By considering the variation of the energy functional for a small displacement of the centre of mass of the droplets in the axial direction, we find that the frequency of this out-of-phase oscillation is given by $\omega_x\sqrt{1-\frac{2}{m\omega_{xN}^2}\int \mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2V_{\mathrm{dd}}(\mathbf{r}_1-\mathbf{r}_2)\frac{\partial n_1(\mathbf{r}_1)}{\partial x}\frac{\partial n_2(\mathbf{r}_2)}{\partial x}}$ (ref. [41]). This integral can easily be evaluated numerically and, for the choice $\epsilon_{\mathrm{dd}}=1.53$, gives the result $\omega/\omega_x=1.95$, in relatively good agreement with the value $\omega/\omega_x=1.85$ obtained from real-time simulations.

**Experimental methods.** The experiments begin with a BEC of $^{162}$Dy atoms, with typical atom number $N=3.5\times10^4$, with no detectable thermal fraction, in a crossed-beam dipole trap (see ref. [42] for details). Typical trap frequencies are $\omega_x$, $\omega_y$ and $\omega_z=2\pi\times(18.5,\ 53,\ 81)$ Hz. Since there are day-to-day variations of the order of 5%, the trap frequencies are measured after each oscillation experiment.

The contact scattering length $a_{\mathrm{s}}$ is controlled using a set of Feshbach resonances; we use the precise $a_{\mathrm{s}}(B)$ conversion provided in ref. [43]. We calibrate the magnetic-field amplitude through radio-frequency spectroscopy between two hyperfine states, with uncertainty of about 1 mG (ref. [10]). The overall systematic uncertainty on the absolute value of $a_{\mathrm{s}}$ is about $3a_0$, which corresponds to an uncertainty on $\epsilon_{\mathrm{dd}}$ of about 4%. To circumvent it, we identify a precise $B$-to-$a_{\mathrm{s}}$ conversion by comparing the experimental and numerical data for the critical $\epsilon_{\mathrm{dd}}$ for the roton instability. The experimental resolution of $a_{\mathrm{s}}$ is about $3a_0$.We checked, both experimentally and numerically, that the day-to-day variations of $a_{\mathrm{s}}(B)$, of the trap frequencies and of the atom number produce a shift of the critical $\epsilon_{\mathrm{dd}}$ that is smaller than our resolution.

The condensate is initially created at $a_{\mathrm{s}}=140a_0$. The scattering length is then tuned with an 80-ms ramp to $a_{\mathrm{s}}=114a_0$, close to the roton instability, which occurs at $a_{\mathrm{s}}=94a_0$ ($\epsilon_{\mathrm{dd}}\approx1.38$). Depending on the regime studied, we use different methods for exciting the axial breathing mode, as follows. (1) In the supersolid regime, the axial oscillation is naturally triggered by crossing the phase transition slowly from the BEC side, with a 30-ms ramp in $a_{\mathrm{s}}$. The typical oscillation amplitude of about 10% sets the minimum amplitude achievable experimentally in the supersolid regime. (2) In the BEC regime, we can excite a small-amplitude oscillation (about 5%) by changing the scattering length from $a_{\mathrm{s}}=114a_0$ to the final value in 10 ms; see the point at $\epsilon_{\mathrm{dd}}=1.35$ in Fig. 3 and Extended Data Fig. 2c. (3) For exciting large-amplitude oscillations in the BEC regime, we cross a narrow resonance located at approximately 5.5 G; see points below $\epsilon_{\mathrm{dd}}=1.35$ in Fig. 3 and Extended Data Fig. 2e. By changing the duration of the crossing ramp, we can tune the amplitude of the oscillation in the range 20–50%.

After variable waiting time at the final scattering length, we record the atomic distribution in the $x$–$y$ plane by absorption imaging after 95 ms of free expansion. Within 200 μs before the release of the atoms from the trapping potential, we set $a_{\mathrm{s}}=140a_0$, thus minimizing the effects of the dipolar interaction on the expansion[12]. We interpret the recorded distributions as momentum-space densities, $n(k_x,k_y)$. The imaging resolution is 0.2 μm$^{-1}$ (1/e Gaussian width).

**Transverse directions and other oscillation modes.** The axial breathing mode is the lowest compressional mode of the dipolar BEC[34]. We have checked in both theory and experiment that it couples mainly to the width along the $x$ direction, while the oscillations of the $y$ and $z$ widths have negligible amplitudes. The second compressional mode appears in the experiment only for large-amplitude oscillations of the lowest mode, it has a higher frequency and it couples mainly to the $y$ width. For the small-amplitude oscillations studied experimentally in the supersolid regime, a very small oscillation of the $y$ width is barely discernible, with a frequency seemingly close to the one along $x$. For the excitation procedure chosen in the theory, in the supersolid regime higher compressional modes with larger frequencies can also get excited (see Extended Data Fig. 1). These modes couple mainly to the $y$ and $z$ widths and are therefore not relevant for the present analysis. Although the present investigation is focused on the excitation of compressional modes of even parity, additional modes of odd nature could be also excited using alternative approaches.

**Experimental fitting procedures.** For each magnetic field and evolution time, we recorded between 10 and 20 distributions. For each measurement, we determine the one-dimensional distribution $n(k_x)$ by integration along the $y$ direction. We evaluate the second moment of $n(k_x)$ to be $\sigma_k^2=\sum_x(k_x-k_0)^2n(k_x)$. In the supersolid regime, we use a double-slit fitting model[10]:

$$n(k_x)=C_0\mathrm{e}^{-(k_x-k_0)^2/2\sigma^2}[1+C_1\cos^2((k_x-k_0)\pi/\bar{k}+\phi)].$$

The model describes two components: a Gaussian distribution with a periodic modulation of period $\bar{k}$ and amplitude $C_0C_1$, which is essentially the Fourier transform of the in-trap lattice, and an unmodulated Gaussian distribution of amplitude $C_0$, which resembles the measured distribution in $k$ space of an ordinary BEC and is related to the unmodulated component of the in-trap density. Given the complex expansion dynamics of dipolar quantum gases[44], it is not possible to relate analytically the distributions in $k$ space with those in real space. We observe, however, that the measured mean value of $\bar{k}$ agrees with the spatial frequency of the stationary supersolid density found in the theory. The modulation amplitude is defined as $A_k = C_1/(1 + C_1)$ and describes the weight of the modulated component. In the analysis of the supersolid regime, we perform a selection excluding the data that does not show a clear stripe pattern: amplitude $C_1 < 0.35$ and fitting error on $\bar{k} > 10\%$. The selected dataset is then used to provide data for $\sigma_k(t)$, $\bar{k}$, and $A_k$, as shown in Fig. 2.

**Experimental oscillations analysis.** For each magnetic field reported in Fig. 3, we analyse the oscillation dynamics of $\sigma_k$ and, for the supersolid, of $\bar{k}$ and $A_k$. Each observable is fitted with a damped sinusoid:

$$X(t) = \Delta X \ \sin\left(t\sqrt{\omega^2 - \tau^{-2}} + \varphi\right)\mathrm{e}^{-t/\tau} + X_0 + \alpha t$$

where $X(t)$ represents $\sigma_k(t)$, $\bar{k}(t)$ and $A_k(t)$, and $\Delta X, X_0, \omega, \tau, \varphi, \alpha$ are free fitting parameters. In particular, $\Delta X/X_0$ is the relative amplitude of the oscillation, $\omega$ is the frequency and $\tau$ is the damping time. A finite $\alpha$ accounts for possible slow drifts of the observables. For each dataset, we measured the trap frequency $\omega_x$ by exciting a collective dipole oscillation.

A clear oscillation of $\sigma_k(t)$, $\bar{k}(t)$ and $A_k(t)$ is visible only for small-amplitude oscillations and for a narrow range of scattering lengths, $1.38 < \epsilon_{\mathrm{dd}} < 1.45$, corresponding to the supersolid regime. As shown in Extended Data Fig. 2a, for $\epsilon_{\mathrm{dd}} = 1.50(5)$, close to the droplet regime, it is not possible to define $\bar{k}(t)$ and $A_k(t)$, because three-body recombination profoundly affects the dynamics of the system: dense droplets form and rapidly evaporate, destroying the lattice periodicity[10]. This corresponds to a fast loss in the atom number, which we measure experimentally. Extended Data Fig. 2c and d shows two examples of small-amplitude oscillation of $\sigma_k(t)$. In the BEC regime, $\epsilon_{\mathrm{dd}} = 1.35(3)$, $\sigma_k(t)$ oscillates as expected for a hydrodynamic superfluid: the value of the frequency $\omega/\omega_x = 1.48(10)$ is consistent with theory; see Fig. 3. Close to the droplet regime, $\epsilon_{\mathrm{dd}} = 1.50(5)$, $\sigma_k(t)$ oscillates with higher frequency and $\omega/\omega_x = 1.71(14)$.

In the BEC regime, we can precisely measure the mode frequency by exciting large-amplitude oscillations, which allow several periods to be observed. Extended Data Fig. 2e, f shows two examples of large-amplitude oscillations. In the BEC regime, $\epsilon_{\mathrm{dd}} = 1.21(2)$, the oscillation frequency is $\omega/\omega_x = 1.46(2)$, in agreement with the theory. In the supersolid regime, $\epsilon_{\mathrm{dd}} = 1.38(2)$, large-amplitude oscillations do not allow us to distinguish clear oscillations of $\bar{k}(t)$ and $A_k(t)$. However, as shown in Extended Data Fig. 2f, we observe a frequency shift in the oscillation of $\sigma_k(t)$ from short times to long times, consistent with a decay of the system from the supersolid regime to the BEC regime in the first 150 ms, owing to atom losses.

**Large temperature measurements in the BEC regime.** Since finite temperatures can affect the axial breathing mode, we have performed a series of measurements in the BEC regime ($\epsilon_{\mathrm{dd}} = 1.14(2)$), where we can define an equilibrium temperature $T$. To obtain large temperatures, we interrupt the evaporative cooling procedure. The results are summarized in Extended Data Fig. 3. We observe a small negative shift of the oscillation frequency for increasing $T$. Close to the critical BEC temperature, the shift from the measurements at the lowest temperature is $\Delta\omega/\omega \approx -10\%$. We also observe an increase of the damping rate for increasing $T$. Both phenomena can be justified as effects of the interaction of the condensate with the thermal component[45]. For a thermal gas above the condensation temperature, we also observe the axial breathing mode, but with a completely different frequency, $\omega/\omega_x = 2.02(3)$, which is compatible with that expected for the breathing oscillation of a weakly interacting classical gas, $\omega/\omega_x = 2$. These measurements demonstrate that our system is in the collisionless regime, so that the peculiar frequency observed for the axial breathing mode of the condensed component must be attributed to the locking of the phase of the wavefunction[25]. Our knowledge of the lowest temperature achieved in the system is limited, since the temperature can no longer be measured when the thermal fraction drops below approximately 30%. Since the damping time decreases with $T$, we took care to maximize the damping time to 150–200 ms, in both the BEC and supersolid regimes.

## Data availability

The datasets that support the findings of this study are available from the corresponding authors upon reasonable request.

40. Menotti, C. & Stringari, S. Collective oscillations of a one-dimensional trapped Bose-Einstein gas. *Phys. Rev. A* **66**, 043610 (2002).
41. Matveeva, N., Recati, A. & Stringari, S. Dipolar drag in bilayer harmonically trapped gases. *Eur. Phys. J. D* **65**, 219–222 (2011).
42. Lucioni, E. et al. Dysprosium dipolar Bose-Einstein condensate with broad Feshbach resonances. *Phys. Rev. A* **97**, 060701 (2018).
43. Böttcher, F. et al. Quantum correlations in dilute dipolar quantum droplets beyond the extended Gross-Pitaevskii equation. Preprint at https://arxiv.org/abs/1904.10349 (2019).
44. Giovanazzi, S. et al. Expansion dynamics of a dipolar Bose-Einstein condensate. *Phys. Rev. A* **74**, 013621 (2006).
45. Giorgini, S. Collisionless dynamics of dilute Bose gases: Role of quantum and thermal fluctuations. *Phys. Rev. A* **61**, 063615 (2000).

**Acknowledgements** This work received funding from the EC-H2020 research and innovation program (grant number 641122-QUIC). We acknowledge discussions with R. Citro and J. G. Maloberti and technical assistance from A. Barbini, F. Pardini, M. Tagliaferri and M. Voliani. S.M.R., A.R. and S.S. acknowledge funding from Provincia Autonoma di Trento and the Q@TN initiative. We acknowledge discussions with the participants of the Stuttgart meeting on ‘Perspectives for supersolidity in dipolar droplet arrays’.

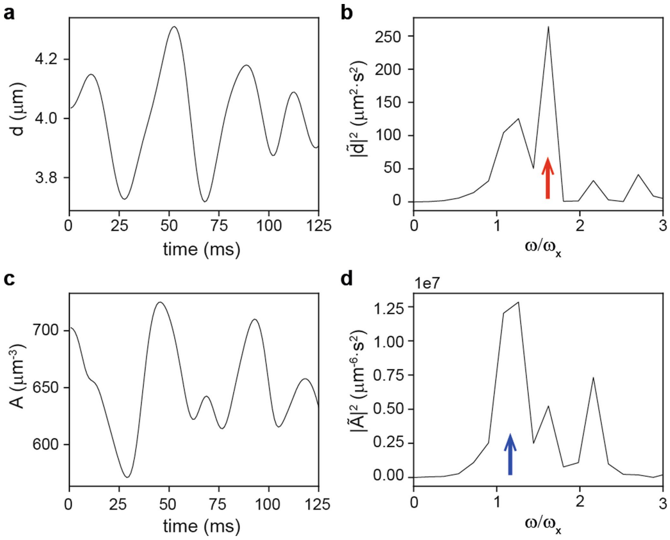
**Author contributions** L.T., E.L., F.F., A.F., C.G. and G.M. conceived the experimental investigation, performed the measurements and the experimental data analysis. S.M.R., A.R. and S.S. conceived the theoretical investigation, performed the simulations and the theoretical data analysis. All authors contributed to discussions and writing of the paper.

**Competing interests** The authors declare no competing interests.

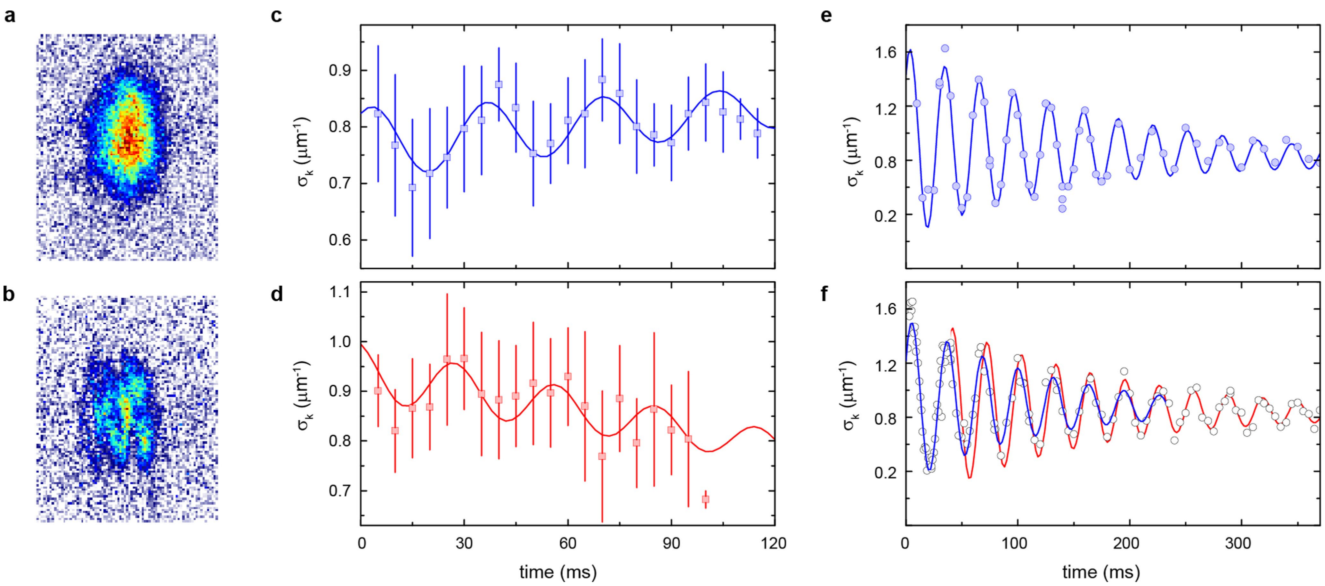
**Additional information**

**Correspondence and requests for materials** should be addressed to G.M. or A.R.

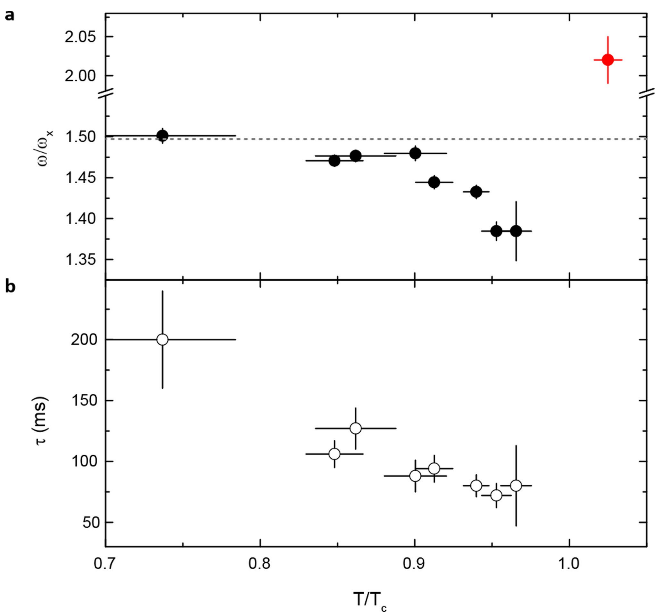
**Peer review information** *Nature* thanks Sean Mossman, Georgy Shlyapnikov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



**Extended Data Fig. 1 | Peak amplitude and relative distance.** In the supersolid regime, the axial breathing mode bifurcates into higher- and lower-frequency modes, mainly coupled to the relative distance between the density peaks and to their amplitude, respectively. **a**, **b**, Time evolution of the relative distance $d(t)$ between two density peaks (**a**) and its Fourier transform, dominated by a peak at the higher frequency (red arrow) (**b**). **c**, **d**, Time evolution of the peak density amplitude $A(t)$ (**c**) and its Fourier transform, dominated by the lower-frequency mode (blue arrow) (**d**).



**Extended Data Fig. 2 | Examples of small- and large-amplitude oscillations.** **a**, **b**, Typical false-colour experimental distributions in the $\left(k_x, k_y\right)$ plane for the BEC regime close to the roton instability, $\epsilon_{\mathrm{dd}} = 1.35(3)$ (**a**) and the droplet regime, $\epsilon_{\mathrm{dd}} = 1.50(5)$ (**b**). **c**, **d**, Small-amplitude oscillation of $\sigma_k(t)$ under the same conditions, $\epsilon_{\mathrm{dd}} = 1.35(3)$ (**c**) and $\epsilon_{\mathrm{dd}} = 1.50(5)$ (**d**). **e**, **f**, Large-amplitude oscillation of $\sigma_k(t)$ for the BEC regime, $\epsilon_{\mathrm{dd}} = 1.21(2)$ (**e**) and the supersolid regime, $\epsilon_{\mathrm{dd}} = 1.38(2)$ (**f**). In the supersolid regime, we observe a small frequency shift in the oscillation of $\sigma_k(t)$ over time: we fit $\omega/\omega_x = 1.43(1)$ up to about 150 ms (blue line), and $\omega/\omega_x = 1.48(2)$ from 150 ms onwards (red line). Error bars represent the standard deviation of 4–8 measurements.



**Extended Data Fig. 3 | Large temperature measurements in the BEC regime. a**, **b**, Measured frequency (**a**) and measured damping time (**b**) of the axial breathing mode at $\epsilon_{dd} \approx 1.2$ for increasing temperature. The measured frequency for a thermal gas (red dot) is compatible with $\omega/\omega_x = 2$, demonstrating the collisionless nature of the system and the breaking of phase invariance for the BEC. Error bars represent the standard deviation of 4–8 measurements.

LETTER

https://doi.org/10.1038/s41586-019-1569-5

# The low-energy Goldstone mode in a trapped dipolar supersolid

Mingyang Guo[1,2,4], Fabian Böttcher[1,2,4], Jens Hertkorn[1,4], Jan-Niklas Schmidt[1,2], Matthias Wenzel[1,2], Hans Peter Büchler[2,3], Tim Langen[1,2] & Tilman Pfau[1,2]*

**A supersolid is a counter-intuitive state of matter that combines the frictionless flow of a superfluid with the crystal-like periodic density modulation of a solid[1,2]. Since the first prediction[3] in the 1950s, experimental efforts to realize this state have focused mainly on helium, in which supersolidity remains unobserved[4]. Recently, supersolidity has also been studied in ultracold quantum gases, and some of its defining properties have been induced in spin–orbit-coupled Bose–Einstein condensates (BECs)[5,6] and BECs coupled to two crossed optical cavities[7,8]. However, no propagating phonon modes have been observed in either system. Recently, two of the three hallmark properties of a supersolid—periodic density modulation and simultaneous global phase coherence—have been observed in arrays of dipolar quantum droplets[9–11], where the crystallization happens in a self-organized manner owing to intrinsic interactions. Here we directly observe the low-energy Goldstone mode, revealing the phase rigidity of the system and thus proving that these droplet arrays are truly supersolid. The dynamics of this mode is reminiscent of the effect of second sound in other superfluid systems[12,13] and features an out-of-phase oscillation of the crystal array and the superfluid density. This mode exists only as a result of the phase rigidity of the experimentally realized state, and therefore confirms the superfluidity of the supersolid.**

Symmetry breaking is a crucial concept for describing phase transitions in particle[14] and condensed matter physics[15,16]. A spontaneous symmetry breaking occurs when the Hamiltonian of a system is invariant with respect to a certain symmetry while the equilibrium ground state is not. An additional order parameter is therefore necessary to describe the system. A broken continuous symmetry, such as translational invariance or the $U(1)$ symmetry associated with particle number conservation, leads to two types of collective excitation called the Goldstone[17] and Higgs[18] modes. These modes can usually be identified from the resulting effective potential, which has the shape of a Mexican hat when plotted with respect to the order parameter. A schematic example is shown in Fig. 1a. In this effective potential the gapless Goldstone[17] mode and the gapped Higgs[18] mode correspond to the phase and amplitude modulation of the complex order parameter at long wavelengths, respectively. In Fig. 1b, c, we schematically show the Goldstone and Higgs modes after breaking the continuous translational symmetry for an infinite system, in which they correspond to a spatial shift (phonons) and an amplitude oscillation of the periodic density modulation. Both Goldstone and Higgs modes have successfully been observed using spectroscopic methods in various platforms including superfluid helium[19], solid-state systems[20–22] and ultracold quantum gases[7,23–27].

A quantum mechanical ground state that simultaneously breaks the global $U(1)$ symmetry and the continuous translational symmetry is the long-sought supersolid. A promising system in which to realize supersolidity is a dipolar quantum gas, which inherently features a preferred length scale for a periodic density modulation owing to its roton-like dispersion relation[28–30]. This roton wavelength can be tuned through the external trapping potential[29]. For increasingly dominant dipolar interactions, the roton mode softens and finally becomes imaginary, at which point mean-field theory predicts a collapse. However, this collapse is prevented by beyond-mean-field effects, leading to stable quantum droplets. Supersolid ground states formed by arrays of these droplets have been proposed for suitably chosen combinations of confinement and scattering length[31,32]. Following this, arrays of quantum droplets featuring supersolid properties—specifically, the crystalline density modulation and global phase coherence—have recently been realized experimentally[9–11]. Simply by tuning the contact interaction, the system undergoes a phase transition from a normal BEC to a supersolid, and then a crossover to an array of isolated droplets. Like a normal crystal, the isolated droplet array has a low compressibility. The coherent array, however, maintains a higher compressibility owing to the compressible BEC component. Distinct from spin–orbit-coupled BECs[5] and BECs coupled to two crossed optical cavities[7,8], spatial ordering of these droplet arrays arises from intrinsic interactions between the particles, and therefore the crystal can support phonon modes. However, a definite proof of superfluid phase rigidity, a synonym for superfluid stiffness, is still missing and is required in order to verify the supersolid nature of the system.

Here we describe the experimental realization of a state that simultaneously shows the three necessary hallmarks of a supersolid—a periodic density modulation, global phase coherence and phase rigidity. In our work, the phase rigidity is proved by studying the low-energy Goldstone mode of the system, which arises from the two broken symmetries. The observed low-energy Goldstone mode features an out-of-phase oscillation between the droplet array and the superfluid density, involving Josephson-like dynamics between the droplets and therefore highlighting the phase rigidity of the state. The out-of-phase Goldstone mode emerges precisely in the supersolid region and provides compelling evidence of the presence of supersolidity.

For an infinite supersolid droplet array that simultaneously breaks the global $U(1)$ symmetry and the continuous translational symmetry, we can define the corresponding two complex order parameters as the condensate wavefunction $\Psi(k=0)$ and the periodic density modulation $\varrho\left(k=\frac{2\pi}{d}\right)=\mathcal{F}[|\Psi(x)|^2]$ with a wavelength $d$. Accompanying the two order parameters, there are two branches of Goldstone and Higgs modes. This leads to a complex excitation spectrum[33,34], featuring two kinds of sound, one corresponding to the compressibility of the supersolid and one corresponding to the superfluid stiffness. Correspondingly, for a trapped supersolid droplet array with a finite size, the two Goldstone modes consist of an in-phase combination of phonons in the BEC and the crystal part of the supersolid (Fig. 2a), and an out-of-phase oscillation of the two (Fig. 2b). The in-phase mode is exactly the centre-of-mass (COM) dipole oscillation of the whole cloud at the trap frequency. However, the out-of-phase mode maintains the COM by a precise interplay of the crystal motion and the superfluid counterflow, which can have a much lower energy. In contrast to the low-energy Goldstone mode, the Higgs mode in our system is expected

---

[1] 5. Physikalisches Institut, Universität Stuttgart, Stuttgart, Germany. [2] Center for Integrated Quantum Science and Technology, Stuttgart, Germany. [3] Institute for Theoretical Physics III, Universität Stuttgart, Stuttgart, Germany. [4] These authors contributed equally: Mingyang Guo, Fabian Böttcher, Jens Hertkorn. *e-mail: t.pfau@physik.uni-stuttgart.de

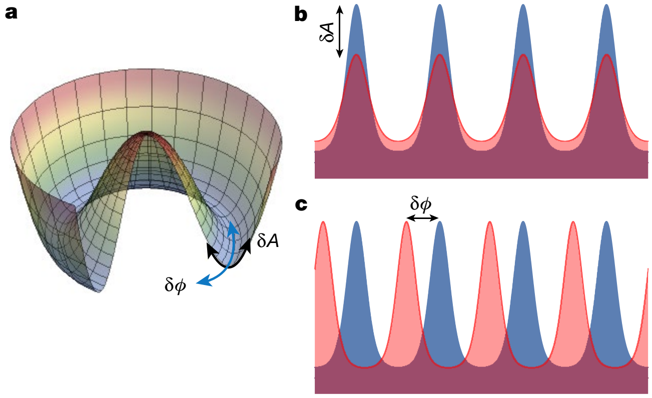


**Fig. 1 | Goldstone and Higgs modes arising from the broken continuous translational symmetry in an infinite supersolid. a**, Effective potential of an ordered phase as a function of the complex order parameter. Higgs and Goldstone modes correspond to amplitude δ$A$ and phase δ$\phi$ modulation of the order parameter $\Psi = Ae^{i\phi}$. The gapless Goldstone mode leads to a degeneracy of the ground state. **b**, **c**, Illustration of the Higgs (**b**) and Goldstone (**c**) modes for a broken continuous translational symmetry in an infinite system, for which the modes correspond to an amplitude modulation and a spatial shift of the crystal structure, respectively.

to be strongly damped and therefore has only a negligible effect on our observations.

To thoroughly explore the two modes in trapped droplet arrays, we implement Bogliubov–de Gennes theory based on the extended Gross–Pitaevskii equation, including the relevant beyond-mean-field corrections. In the supersolid region, both the in-phase and out-of-phase modes are found at the long-wavelength limit that corresponds to the sample size (see Methods). As expected, the energy of the in-phase mode exactly corresponds to the trap frequency, while the out-of-phase Goldstone mode has an excitation energy much lower than the trap frequency. The simulated time evolution of the out-of-phase mode in a supersolid array is shown in Fig. 2c. Starting from the symmetric three-droplet ground state, the droplet array moves to one side. This movement of the droplet array is accompanied by a superfluid atom transport that redistributes the atoms in order to maintain the COM. For large displacements, which are realized for larger excitation amplitudes, we observe an oscillation between a three- and four-droplet state (see Methods). For each time step we confirm that the COM stays unchanged within our simulation resolution. Furthermore, from this simulation we obtain an oscillation frequency of 5.6 Hz for this low-energy mode, using the parameters of the simulation of Fig. 2c, in agreement with the Bogoliubov result.

To maintain the COM, the superfluid flow characterized by the droplet imbalance $\eta$ of the droplet array and the crystal displacement $\Delta x$ must satisfy a certain relation. In an array of three droplets, we define the imbalance as $\eta = (N_1 - N_3)/(N_1 + N_2 + N_3)$, with $N_i$ the atom number in the $i$th droplet, numbered from left to right, and the displacement $\Delta x$ as the arithmetic mean of the positions of the three droplets relative to the COM of the whole cloud, also including a thermal background. Figure 2d shows the strong correlation between $\eta$ and $\Delta x$ extracted from the numerical simulation shown in Fig. 2c, which is found to be the same for different excitation amplitudes. More importantly, the correlation is also robust against a small variation of the scattering length, corresponding to different fractions of the superfluid background. Therefore, the existence of such a correlation acts as a clear signature for the out-of-phase Goldstone mode, and thus proves the supersolidity of the system.

To experimentally observe this low-energy Goldstone mode, we prepared ultracold quantum gases of the bosonic dysprosium isotope $^{162}$Dy in an elongated optical dipole trap with trap frequencies $\omega = 2\pi[30(1), 89(2), 108(2)]$ Hz and the magnetic field along the $y$ axis (see Methods). Depending on the final scattering length, the prepared cloud can be in the normal BEC phase ($>100a_0$), an incoherent droplet array ($<96a_0$), or a coherent droplet array in between ($a_0$ is the Bohr radius). After reaching the final scattering length, we let the cloud evolve and equilibrate for 15 ms before we probe it using in situ phase-contrast imaging.

Because the oscillation period of the out-of-phase mode is much longer than the droplet lifetime, which is limited by three-body losses[9], it is currently not feasible to use a standard spectroscopic method[23,30] to detect the Goldstone mode. However, the low excitation energy of the

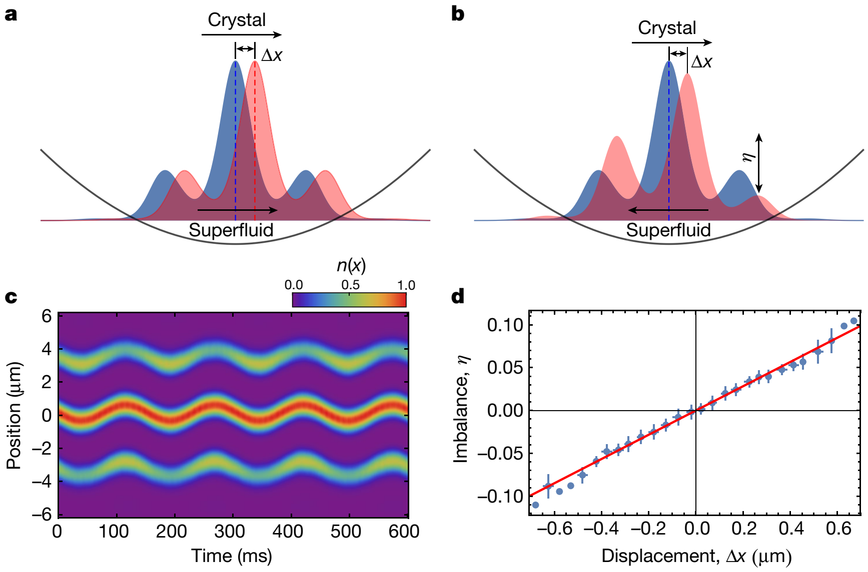


**Fig. 2 | Goldstone modes in a harmonic potential. a**, **b**, The in-phase (**a**) and out-of-phase (**b**) Goldstone modes in a harmonic trap in the long-wavelength limit. The arrows indicate the spatial displacement $\Delta x$ of the droplet array and the flow of the superfluid density. Vertical dashed lines indicate the COM position of the whole cloud. The in-phase mode corresponds to the COM dipole oscillation with the trap frequency, while the out-of-phase mode maintains the COM owing to the counterflow of the droplet array and the superfluid background. As a result, this mode has a very low energy. **c**, Simulated dynamics of the out-of-phase Goldstone mode for the ground state, an array of three droplets. The dynamics is illustrated by the normalized line density along the $x$ axis, $n(x)$. The spatial motion of the droplet array is compensated by a superfluid flow of atoms in the opposite direction in order to maintain the COM. **d**, Numerically predicted correlation between the imbalance $\eta$ and the displacement $\Delta x$ of the droplet array, together with a linear fit. The error bars shown indicate the uncertainty of the fit used to extract $\eta$ and $\Delta x$.

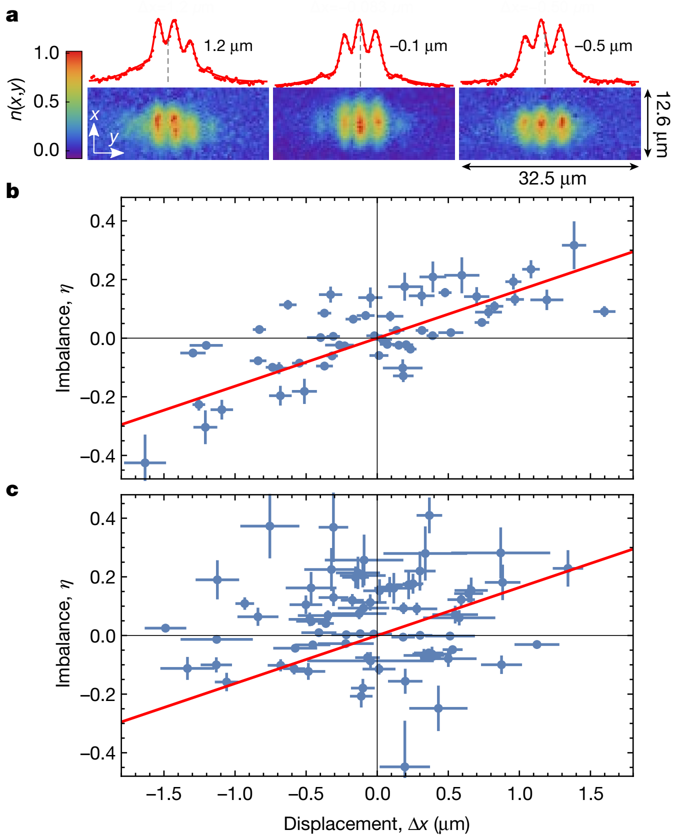


**Fig. 3 | Experimental correlation between $\eta$ and $\Delta x$.** **a**, Exemplary in situ images of the normalized density $n(x,y)$ with different displacements of the droplet array relative to the total COM of the whole cloud (dashed line). When the droplet array shifts to the right (left) with a positive (or negative) $\Delta x$, atoms flow to the left (or right), resulting in higher atom number on the left (or right) side, in order to keep the COM unchanged. The integrated densities together with the fits (red curves) are also shown. **b**, **c**, Experimental correlations in the supersolid ($97.6a_0$) (**b**) and isolated droplet ($91.2a_0$) (**c**) regions, respectively. The error bars correspond to the uncertainty of the fits used to extract the position and atom number of each individual droplet. The red curve is the theoretical prediction for a supersolid state without any free parameters, as shown in Fig. 2d. A correlation is clearly demonstrated in the supersolid region, whereas such a correlation is missing in the isolated droplet phase.

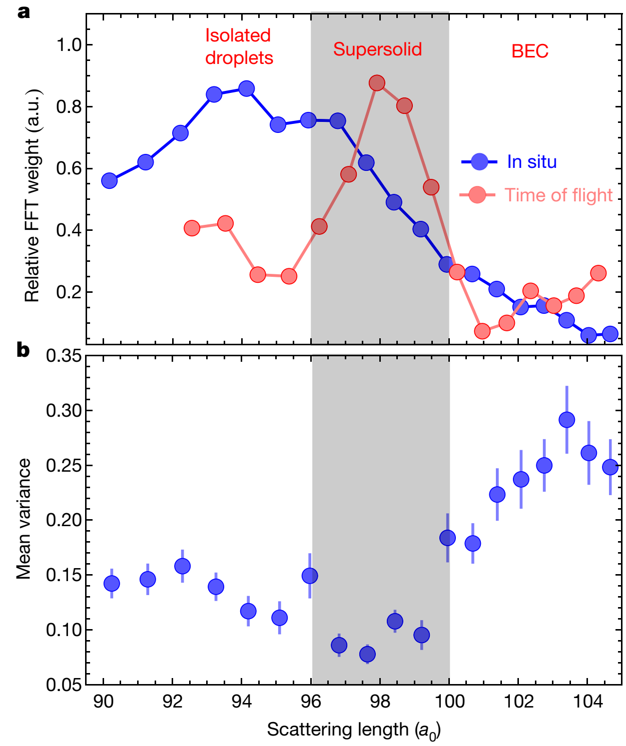


**Fig. 4 | The $\eta$–$\Delta x$ correlation across the phase diagram.** **a**, Experimental signature of the phase diagram for our trap geometry. We can identify the different regions by evaluating the in situ density modulation (blue, characterized by the corresponding Fourier weight at the modulation wavelength) and the global phase coherence (red, characterized by averages of the complex fast Fourier transform (FFT) weights for more than 50 repeated time-of-flight interference patterns)[9]. The shadowed area is the coherent droplet region we determined. a.u., arbitrary units. **b**, Mean variance of the experimental data with respect to the theoretical correlation curve for the supersolid state. Coinciding exactly with the phase-coherent region in **a** is the range where we observe the smallest variance, and therefore the strongest correlation, which indicates the existence of the low-energy Goldstone mode. The errors bars indicate one standard error of the experimental data with respect to the correlation obtained from the theoretical simulations.

out-of-phase Goldstone mode leads to its excitation within the energy bandwidth of the dynamical formation process. As a result, the mode should always be excited in the prepared samples with a different phase owing to experimental imperfections, such as the non-adiabatic ramping of the scattering length, technical noise and thermal fluctuations. We therefore repeated the experiment many times at each scattering length, in order to statistically map out the correlation between the imbalance $\eta$ and the crystal displacement $\Delta x$.

After post-selection on the total atom number (see Methods), $\Delta x$ and $\eta$ are extracted for each cloud by fitting the in situ density distributions with a sum of four Gaussian functions, corresponding to a BEC or thermal background and three droplets. Some example images with different imbalances and displacements $\Delta x$ for a scattering length of $97.6a_0$ are shown in Fig. 3a to visualize how the imbalance changes with respect to the spatial shift of the droplets. Figure 3b shows the measured correlation between $\eta$ and $\Delta x$ for the observed three-droplet states in the supersolid region, where a strong correlation is observed. More importantly, the correlation perfectly coincides with the theoretical prediction (red curve in Fig. 3b) without any free parameters. We interpret the spread in the data around the theoretical correlations as imperfections in the extraction of the imbalance and the displacement, as well as small excitations of higher-lying modes. The existence of the correlation proves the presence of the low-energy Goldstone mode, and therefore the supersolidity of the system. For comparison, one example of isolated droplet arrays at $91.2a_0$ is shown in Fig. 3c, where the correlation is missing, resulting from the lack of superfluid counterflow due to the vanishing superfluid fraction.

To quantify the range of the contact interaction for which we can observe supersolid arrays of quantum droplets, we calculated the variance of the experimental data with respect to the obtained correlation from the theoretical simulations. A small variance is therefore evidence for the existence of the low-energy Goldstone mode. The measured variance of the data compared to the theory across the explored region is shown in Fig. 4b. We can clearly see a region with smaller variance, and therefore stronger correlation, which coincides exactly with the region of global phase coherence (shadowed area in Fig. 4a). The latter can independently be determined from time-of-flight interference[9]. In the BEC phase, some clouds are also detected to be density-modulated, which may result from excitations of the roton mode, which also has a small excitation energy close to the phase transition owing to its softening. However, the correlation is also missing in this region. As a theory-independent test, we analyse the $\eta$–$\Delta x$ correlation data with a linear fit and confirm that only in the supersolid region can the displacement be compensated by a superfluid flow, and therefore an imbalance of the droplet array (see Methods).

In conclusion, we have studied the low-energy Goldstone mode in a trapped supersolid droplet array of dipolar dysprosium atoms, which enables the simultaneous detection of phase coherence and phase rigidity. The out-of-phase mode features a counterflow of the crystal-like droplet array and the superfluid density, leading to a robust correlation between the imbalance and the displacement of the droplets. This mode therefore directly connects to the broken $U(1)$ symmetry and the continuous translational symmetry, highlighting the supersolid nature of the coherent arrays of dipolar quantum droplets. An extension to our work would be the observation of other collective excitations, especially the Higgs modes, which require additional symmetries to suppress their decay into other lower-energy excitations[21]. Another goal could be to realize a supersolid state with larger droplet numbers, or even two-dimensional supersolid arrays[35], where an additional spatial symmetry is broken, leading to an even more complex excitation spectrum.

We note that during the preparation (review process) of this manuscript, we became aware of related complementary studies of higher-lying collective modes[36,37].

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1569-5.

Received: 11 June 2019; Accepted: 29 August 2019;
Published online 9 September 2019.

1. Leggett, A. J. Can a solid be "superfluid"? *Phys. Rev. Lett.* **25**, 1543 (1970).
2. Boninsegni, M. & Prokof'ev, N. V. *Colloquium*: Supersolids: what and where are they? *Rev. Mod. Phys.* **84**, 759 (2012).
3. Gross, E. P. Unified theory of interacting bosons. *Phys. Rev.* **106**, 161 (1957).
4. Chan, M., Hallock, R. & Reatto, L. Overview on solid $^4$He and the issue of supersolidity. *J. Low Temp. Phys.* **172**, 317–363 (2013).
5. Li, J. R. et al. A stripe phase with supersolid properties in spin-orbit-coupled Bose–Einstein condensates. *Nature* **543**, 91–94 (2017).
6. Bersano, T. M. et al. Experimental realization of a long-lived striped Bose-Einstein condensate induced by momentum-space hopping. *Phys. Rev. A* **99**, 051602 (2019).
7. Léonard, J., Morales, A., Zupancic, P., Donner, T. & Esslinger, T. Monitoring and manipulating Higgs and Goldstone modes in a supersolid quantum gas. *Science* **358**, 1415–1418 (2017).
8. Léonard, J., Morales, A., Zupancic, P., Esslinger, T. & Donner, T. Supersolid formation in a quantum gas breaking a continuous translational symmetry. *Nature* **543**, 87–90 (2017).
9. Böttcher, F. et al. Transient supersolid properties in an array of dipolar quantum droplets. *Phys. Rev. X* **9**, 011051 (2019).
10. Tanzi, L. et al. Observation of a dipolar quantum gas with metastable supersolid properties. *Phys. Rev. Lett.* **122**, 130405 (2019).
11. Chomaz, L. et al. Long-lived and transient supersolid behaviors in dipolar quantum gases. *Phys. Rev. X* **9**, 021012 (2019).
12. Atkins, K. R. *Liquid Helium* (Cambridge Univ. Press, 1959).
13. Sidorenkov, L. A. et al. Second sound and the superfluid fraction in a Fermi gas with resonant interactions. *Nature* **498**, 78 (2013).
14. Nambu, Y. & Jona-Lasinio, G. Dynamical model of elementary particles based on an analogy with superconductivity. *Phys. Rev.* **124**, 246 (1961).
15. Griffin, A. et al. *Excitations in a Bose-condensed Liquid* Vol. 4 (Cambridge Univ. Press, 1993).
16. Anderson, P. W. Random-phase approximation in the theory of superconductivity. *Phys. Rev.* **112**, 1900 (1958).
17. Goldstone, J. Field theories with superconductor solutions. *Il Nuovo Cimento (1955–1965)* **19**, 154–164 (1961).
18. Higgs, P. W. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.* **13**, 508 (1964).
19. Yarnell, J., Arnold, G., Bendt, P. & Kerr, E. Energy vs momentum relation for the excitations in liquid helium. *Phys. Rev. Lett.* **1**, 9 (1958).
20. Sooryakumar, R. & Klein, M. Raman scattering by superconducting-gap excitations and their coupling to charge-density waves. *Phys. Rev. Lett.* **45**, 660 (1980).
21. Littlewood, P. & Varma, C. Gauge-invariant theory of the dynamical interaction of charge density waves and superconductivity. *Phys. Rev. Lett.* **47**, 811 (1981).
22. Rüegg, C. et al. Quantum magnets under pressure: controlling elementary excitations in $TlCuCl_3$. *Phys. Rev. Lett.* **100**, 205701 (2008).
23. Stamper-Kurn, D. et al. Excitation of phonons in a Bose–Einstein condensate by light scattering. *Phys. Rev. Lett.* **83**, 2876 (1999).
24. Bissbort, U. et al. Detecting the amplitude mode of strongly interacting lattice bosons by Bragg scattering. *Phys. Rev. Lett.* **106**, 205303 (2011).
25. Endres, M. et al. The 'Higgs' amplitude mode at the two-dimensional superfluid/Mott insulator transition. *Nature* **487**, 454–458 (2012).
26. Hoinka, S. et al. Goldstone mode and pair-breaking excitations in atomic Fermi superfluids. *Nat. Phys.* **13**, 943–946 (2017).
27. Behrle, A. et al. Higgs mode in a strongly interacting fermionic superfluid. *Nat. Phys.* **14**, 781–785 (2018).
28. Santos, L., Shlyapnikov, G. V. & Lewenstein, M. Roton-maxon spectrum and stability of trapped dipolar Bose–Einstein condensates. *Phys. Rev. Lett.* **90**, 250403 (2003).
29. Chomaz, L. et al. Observation of roton mode population in a dipolar quantum gas. *Nat. Phys.* **14**, 442–446 (2018).
30. Petter, D. et al. Probing the roton excitation spectrum of a stable dipolar Bose gas. *Phys. Rev. Lett.* **122**, 183401 (2019).
31. Roccuzzo, S. M. & Ancilotto, F. Supersolid behavior of a dipolar Bose-Einstein condensate confined in a tube. *Phys. Rev. A* **99**, 041601 (2019).
32. Wenzel, M., Böttcher, F., Langen, T., Ferrier-Barbut, I. & Pfau, T. Striped states in a many-body system of tilted dipoles. *Phys. Rev. A* **96**, 053630 (2017).
33. Saccani, S., Moroni, S. & Boninsegni, M. Excitation spectrum of a supersolid. *Phys. Rev. Lett.* **108**, 175301 (2012).
34. Macrì, T., Maucher, F., Cinti, F. & Pohl, T. Elementary excitations of ultracold soft-core bosons across the superfluid-supersolid phase transition. *Phys. Rev. A* **87**, 061602 (2013).
35. Baillie, D. & Blakie, P. Droplet crystal ground states of a dipolar Bose gas. *Phys. Rev. Lett.* **121**, 195301 (2018).
36. Tanzi, L. et al. Supersolid symmetry breaking from compressional oscillations in a dipolar quantum gas. *Nature* https://doi.org/10.1038/s41586-019-1568-6 (2019).
37. Natale, G. et al. Excitation spectrum of a trapped dipolar supersolid and its experimental evidence. *Phys. Rev. Lett.* **123**, 050402 (2019).

# METHODS

**Sample preparation and experimental details.** The complete experimental sequence is described in detail in our previous publications[9,38]. In short, we prepare a quasipure BEC of $^{162}$Dy in a crossed optical dipole trap formed by two laser beams at 1,064 nm. The degenerate cloud typically contains $4 \times 10^4$ atoms at a temperature below 10 nK. After evaporation, the trap is reshaped within 20 ms to the final trap geometry with trap frequencies of $\omega = 2\pi[30(1), 89(2), 108(2)]$ Hz. For the measurements presented, the magnetic field is orientated along the $y$ direction. Subsequently, we tune the contact interaction from the background scattering length $a_{\mathrm{bg}} = 140(20)a_0$ (refs. [39–41]) to approximately $112a_0$ by ramping the magnetic field closer to the double Feshbach resonances near 5.1 G (refs. [9,42,43]). To reach the droplet region, the magnetic field is further linearly ramped to the final scattering length in the range between $90a_0$ and $105a_0$ in 30 ms. We then let the samples evolve for 15 ms to allow the quantum droplet arrays to form and equilibrate. Subsequently we probe the atomic clouds with our in situ phase-contrast imaging, which is performed along the $z$ axis using a microscope objective with a numerical aperture of 0.3. Our resolution of about 1 μm allows us to distinguish nearby droplets that are separated by about 3 μm. By fitting each individual droplet with a Gaussian, we can extract the atom number in each individual droplet, as well as the position of the droplet to a higher precision than our imaging resolution.

To verify the range for which we observe phase-coherent droplet arrays, we implement a time-of-flight interference sequence similar to our previous work[9]. For this we ramp up the scattering length within 100 μs to about $1a_{\mathrm{bg}}$, to accelerate the expansion and then release the atoms from the trap. Owing to the changed geometry the expansion time is now limited to 7.2 ms compared to our previous publication[9]. Nonetheless, we can observe clear interference patterns and are able to distinguish between phase-coherent and incoherent droplet regimes.

Notably, for samples with smaller scattering lengths, which means for magnetic fields closer to the Feshbach resonances, the clouds suffer from more severe three-body loss. The black data points in Extended Data Fig. 1 indicate the corresponding average atom number in the experiment for each scattering length. For our evaluation we post-select realizations in a range ±15% around the average atom number at each scattering length in the statistical evaluation from which the correlation is extracted. The experimental realizations mostly consist of three or four droplets. We analyse these two cases separately after distinguishing them with multiple Gaussian fits. Although the COM is experimentally determined by the whole image, it is dominated by the condensate and thermal background within the BEC and coherent droplet region. For isolated droplets we also observe a thermal background, which then acts as the main contribution to the determined COM of the whole cloud that we use as a reference for the calculation of the displacement.

For the range of scattering lengths studied in this work, the experimentally determined stability of the magnetic field leads to an uncertainty of about $1a_0$, while the calibration of the positions and widths of the double Feshbach resonances, which we use to tune the scattering length, results in an uncertainty of about $4a_0$. On top of this uncertainty, there is an overall systematic uncertainty due to the absolute value of the background scattering length[39–41], which has so far not been measured to high precision. This leads to an uncertainty of all calculated scattering lengths that is of the order of 15%.

**Simulation details.** Our theory is based on numerically solving the extended Gross–Pitaevskii equation[31,44,45], including quantum fluctuations as beyond-mean-field corrections, which prevents the system from collapse and leads to stable quantum droplets[32,38,46–48]. We obtain the ground states by performing imaginary time evolution[32] of the extended Gross–Pitaevskii equation including a harmonic potential with $\omega = 2\pi[30, 90, 110]$ Hz, similar to the experiment. In this trap we obtain a three-droplet ground state for scattering lengths $a_s \geq 91a_0$, as shown in Extended Data Fig. 3. To identify the phase-coherent droplet arrays, we use the same indicator of the nearest minimum divided by the central maximum as an estimation of the overlap between the droplets, as in our previous publication[9]. The result is shown in Extended Data Fig. 1, where we can successfully identify the three different regions, with the coherent region being located at $94a_0$–$97a_0$. Compared to the experimental data, the coherent region obtained from the simulations is shifted by $3a_0$, in line with recent results with erbium atoms showing a similar deviation[11,30].

To calculate the excitation spectrum, we follow the Bogoliubov–de Gennes theory by linearizing the extended Gross–Pitaevskii equation around the ground states[44,45]. The resulting Bogoliubov–de Gennes equations are then solved numerically to obtain the excitation modes[44]. Examples of the dispersion relations obtained are shown in Extended Data Fig. 2 for three different scattering lengths. Owing to finite sample size, the excitation spectra become discrete, with each mode corresponding to a spread out momentum. To visualize the spectra we calculate the zero-temperature structure factor[49,50] $S(\omega, q)$ that indicates the dynamic response of the cloud. The higher the amplitude of this structure factor, the stronger the density response of the system to the corresponding mode. In this work, we ignore the modes with excitation energies higher than about 90 Hz, at which point excitations along the other trap axes start to have a role, making the spectrum much more complex.

In the BEC phase (Extended Data Fig. 2a), there is only one excitation branch, featuring an increasing excitation energy with increasing momentum $q$. By decreasing the contact interaction strength we can observe the appearance of the roton mode in Extended Data Fig. 2b. Lowering the scattering length further, the roton mode softens and we enter the supersolid regime, where an additional excitation branch with lower energy appears, owing to breaking of the continuous translational symmetry. This low-energy mode corresponds to the out-of-phase Goldstone mode, whose correlation we directly observe in the experiment. Close to the phase transition from BEC to supersolid, the calculations clearly reveal a large energy splitting between this low-energy mode and any other collective mode, especially the in-phase COM mode at the trap frequency. For smaller contact interaction strength we reach the regime of isolated droplets, where the low-energy out-of-phase mode decreases in energy. At the same time, we observe the emergence of a clear periodicity in the excitation spectrum owing to the underlying crystal structure.

**Dynamics of the out-of-phase Goldstone mode.** From the numerically solved Bogoliubov–de Gennes equations, we can also obtain the phase pattern of each excitation mode. In Extended Data Fig. 3 we show the calculated one-dimensional line cut through the three-dimensional phase pattern of the out-of-phase Goldstone mode, together with a density cut through the calculated density profile of the three-droplet ground state. As in Josephson dynamics, the phase gradient is directly proportional to the particle flow. Distinct from the phase pattern of the in-phase dipole mode, which is a constant phase gradient over the cloud, the out-of-phase mode has a stepwise phase pattern, with each step coinciding with a droplet. Whereas the whole BEC background thus experiences a phase gradient with a particular direction, the droplets always experience a gradient in the opposite direction. This leads to the counterflow between the BEC background and the crystal that characterizes this mode. By numerically imprinting the phase pattern of a specific collective excitation onto the ground state, we can directly simulate the excitation dynamics of each individual mode by performing a real-time evolution of the extended Gross–Pitaevskii equation. Doing this for the out-of-phase Goldstone mode, we obtain the dynamical time evolution of this low-energy mode, as shown in Fig. 2c.

We simulate the real-time dynamics of the system for different scattering lengths in the supersolid range. Decreasing the scattering length, we observe an increase of the oscillation period, in line with the results obtained from the calculated dispersion relations. This decrease of the oscillation frequency is accompanied by a decrease in the oscillation amplitude. Looking at the correlation between the imbalance $\eta$ and displacement $\Delta x$ that we use as an indication of the presence of the low-energy Goldstone mode, we observe that it remains unchanged within the uncertainty of our evaluation, even though the oscillation frequency nearly doubles in the range of the scattering length that we studied.

Using a similar procedure, we have also checked that the correlation is independent of the excitation amplitude. This is the case as long as the oscillation amplitude remains sufficiently small. In our simulations for larger oscillation amplitudes, we can observe that the three-droplet state changes to a four-droplet state for large displacements. As an example, we show two simulated real-time evolutions with higher amplitudes of the low-energy mode in Extended Data Fig. 4. The change from a three-droplet state to a four-droplet state can happen because in our particular trap and at the studied scattering lengths, the energy difference between the two states is small. However, the periodic coherent emergence and disappearance of a new droplet on the edges of the system is further evidence of the presence of superfluid flow. For very large amplitudes, as shown in Extended Data Fig. 4b, the excitation is no longer oscillatory, but travelling in one direction, reminiscent of the Goldstone mode in an infinite system (see Fig. 1c).

**Evaluation of four-droplet states.** For large displacements in our dynamical simulations of the Goldstone mode, which happen at a large amplitude of the excitation, the three-droplet ground state can change to a four-droplet state. Since the two states with different droplet number are smoothly connected, we also observe four-droplet states in the experiment, owing to strong excitations of the Goldstone mode.

To maintain the COM, this mode should again satisfy a certain relation between the imbalance, which we now define as $\eta = (N_1 + N_2 - N_3 - N_4)/(N_1 + N_2 + N_3 + N_4)$ for a four-droplet state, and the displacement $\Delta x$, which we again define as the arithmetic mean of all four positions of the individual droplets relative to the COM of the whole cloud. The calculated correlation of a four-droplet state is shown in Extended Data Fig. 5. The data in this plot correspond to the large displacement data that are shown in Extended Data Fig. 4. Again, we obtain a linear dependence that is robust against variations of the scattering length, as well as the excitation amplitude.

In our experiment we also find a sizeable amount of realizations with four droplets, for which we perform the same statistical analysis as we presented in the main text for the three-droplet states. As for the three-droplet state, we find a clear correlation for the supersolid regime (Extended Data Fig. 6a), while the variance of the data are larger for the isolated droplet regime (Extended Data Fig. 6b), indicating that no correlation exists in the isolated droplet region. This can also be seen in the example images of four-droplet states in the supersolid regime with different displacements, which are shown in Extended Data Fig. 6c. Similar to the three-droplet case, we can again calculate the variance of the experimental data with respect to the theoretical correlation curve as an indicator of the existence of the Goldstone mode with respect to the contact interaction strength. Similar to Fig. 4b, the variance of the four-droplet state shown in Extended Data Fig. 7 has a clear minimum in the range where we observe phase-coherent droplet arrays.

**Theory-independent evaluation.** As a test that is independent of our underlying theory, we implement a linear fit to the experimental data at each scattering length and compare the data to this fit, instead of the theoretically obtained correlation. This linear fit allows us to do two complementary tests. First, we can again calculate the spread of our experimental measurements with respect to the linear fit. In agreement with the comparison to the theoretical curve, the calculated variance with respect to the linear fit is again lowest directly in the phase-coherent regime and increases by changing the scattering length in either direction. Second, we can look at the absolute value of the intersection point of the linear fit with the axis of the displacement. For a supersolid state, this intersection should be at zero, corresponding to a symmetric droplet array with a vanishing imbalance at no displacement. For isolated droplets, on the other hand, initial fluctuations during the formation process cannot be compensated, meaning that we can get imbalanced droplet arrays, even if the array is in the centre of the cloud. The intersection points obtained for the observed three- and four-droplet states are shown in Extended Data Fig. 8a, b, respectively. This shows that the intersection point is close to zero only in the supersolid region. Both of these theory-independent checks are additional proof of the existence of a correlation between the imbalance and the displacement arising due from the low-energy Goldstone mode of the system.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

38. Kadau, H. et al. Observing the Rosensweig instability of a quantum ferrofluid. *Nature* **530**, 194–197 (2016).
39. Tang, Y., Sykes, A., Burdick, N. Q., Bohn, J. L. & Lev, B. L. s-wave scattering lengths of the strongly dipolar bosons $^{162}$Dy and $^{164}$Dy. *Phys. Rev. A* **92**, 022703 (2015).
40. Tang, Y. et al. Anisotropic expansion of a thermal dipolar Bose gas. *Phys. Rev. Lett.* **117**, 155301 (2016).
41. Tang, Y. et al. Thermalization near integrability in a dipolar quantum Newton's cradle. *Phys. Rev. X* **8**, 021030 (2018).
42. Baumann, K., Burdick, N. Q., Lu, M. & Lev, B. L. Observation of low-field Fano-Feshbach resonances in ultracold gases of dysprosium. *Phys. Rev. A* **89**, 020701 (2014).
43. Böttcher, F. et al. Quantum correlations in dilute dipolar quantum droplets beyond the extended Gross-Pitaevskii equation. Preprint at https://arxiv.org/abs/1904.10349 (2019).
44. Ronen, S., Bortolotti, D. C. & Bohn, J. L. Bogoliubov modes of a dipolar condensate in a cylindrical trap. *Phys. Rev. A* **74**, 013623 (2006).
45. Wilson, R. M., Ronen, S. & Bohn, J. L. Critical superfluid velocity in a trapped dipolar gas. *Phys. Rev. Lett.* **104**, 094501 (2010).
46. Ferrier-Barbut, I., Kadau, H., Schmitt, M., Wenzel, M. & Pfau, T. Observation of quantum droplets in a strongly dipolar Bose gas. *Phys. Rev. Lett.* **116**, 215301 (2016).
47. Ferrier-Barbut, I., Schmitt, M., Wenzel, M., Kadau, H. & Pfau, T. Liquid quantum droplets of ultracold magnetic atoms. *J. Phys. B* **49**, 214004 (2016).
48. Ferrier-Barbut, I., Wenzel, M., Schmitt, M., Böttcher, F. & Pfau, T. Onset of a modulational instability in trapped dipolar Bose-Einstein condensates. *Phys. Rev. A* **97**, 011604 (2018).
49. Brunello, A., Dalfovo, F., Pitaevskii, L., Stringari, S. & Zambelli, F. Momentum transferred to a trapped Bose-Einstein condensate by stimulated light scattering. *Phys. Rev. A* **64**, 063614 (2001).
50. Blakie, P. B., Ballagh, R. J. & Gardiner, C. W. Theory of coherent Bragg spectroscopy of a trapped Bose-Einstein condensate. *Phys. Rev. A* **65**, 033602 (2002).

**Acknowledgements** We acknowledge discussions with the groups of F. Ferlaino, G. Modugno, L. Santos and T. Pohl, as well as with A. Pelster and A. Balaz. This work is supported by the German Research Foundation (DFG) within FOR2247 under Pf381/16-1 and Bu2247/1, Pf381/20-1 and FUGG INST41/1056-1. T.L. acknowledges support from the EU within Horizon2020 Marie Skłodowska Curie IF (grant numbers 746525 coolDips).

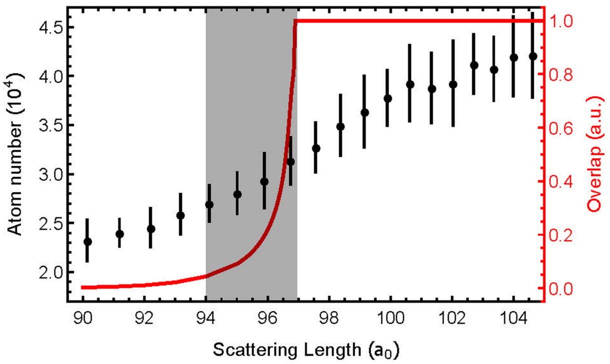
**Author contributions** M.G., F.B. and J.-N.S. performed the experiment and analysed the data. J.H. and M.W. performed the numerical analysis. H.P.B., T.L. and T.P. provided scientific guidance in experimental and theoretical questions. All authors contributed to the interpretation of the data and the writing of the manuscript.

**Competing interests** The authors declare no competing interests.

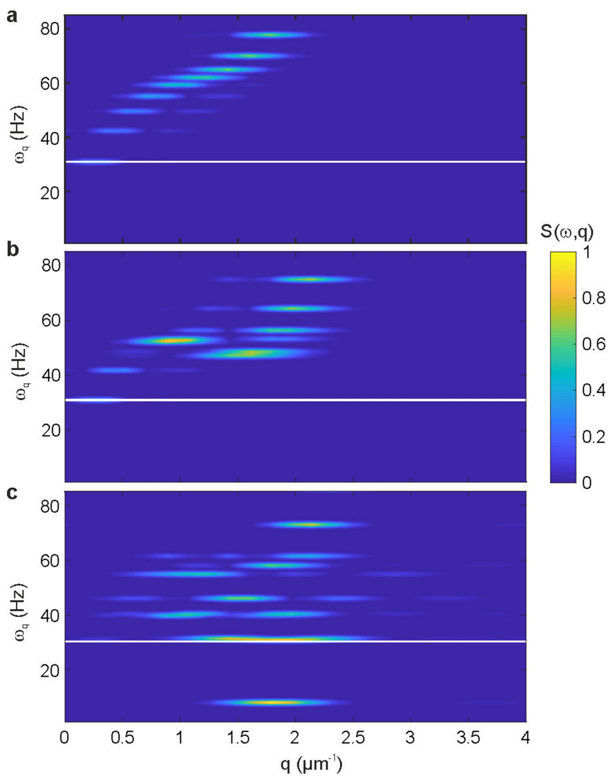
**Additional information**

**Correspondence and requests for materials** should be addressed to T.P.

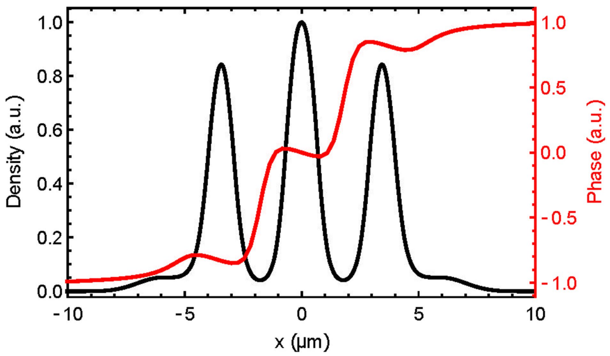
**Peer review information** *Nature* thanks Sean Mossman, Georgy Shlyapnikov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



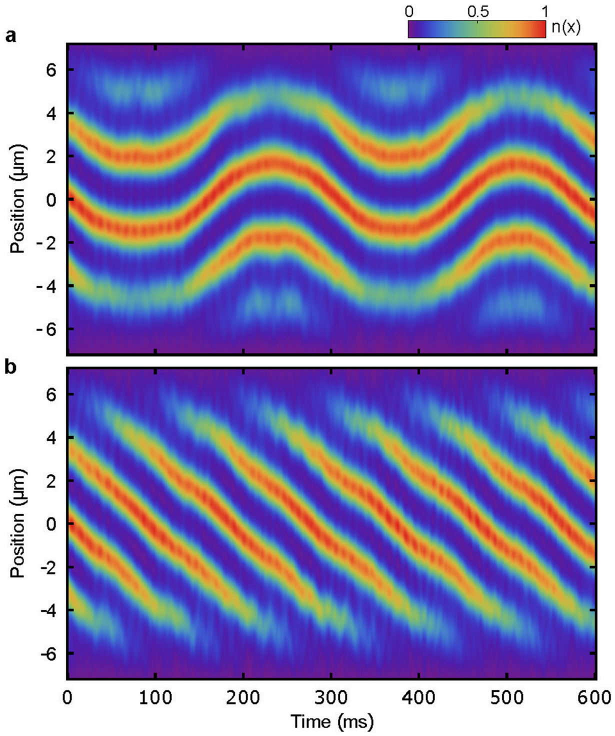
**Extended Data Fig. 1 | Theoretical phase boundaries.** As an indicator of the phase coherence, we show the ratio of the first minimum in the density compared to the central droplet peak density as an indicator of the overlap between the droplets[43] of the calculated density profile of the ground state in red. The shaded area is the coherent region determined from the overlap. As in the experiment, three regions are identified and the coherent region locates between $94a_0$ and $97a_0$, shifting approximately $3a_0$ from the experimentally obtained phase diagram. These simulations are done for an atom number of $30 \times 10^3$. The black points indicate the measured average atom number in the experiment, with the error bars representing standard deviations for more than ten shots at each scattering length. (a.u., arbitrary units.).



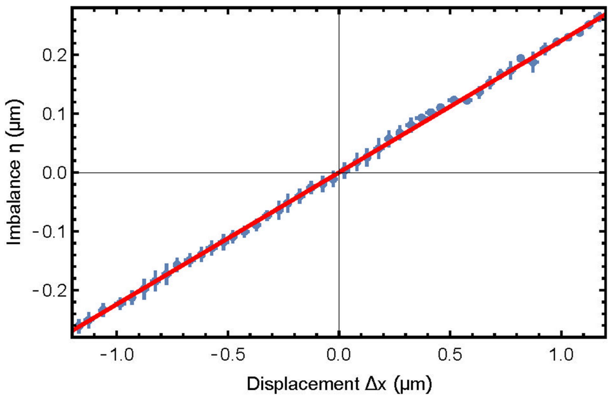
**Extended Data Fig. 2 | Dynamic structure factor of the collective excitations.** Calculated structure factor $S(\omega, q)$ in the BEC phase for the scattering length $a_s = 100a_0$ (**a**) and $a_s = 98a_0$ (**b**) and in the supersolid droplet arrays (**c**) for $a_s = 96a_0$. Owing to the finite size of the system, long-wavelength modes become discrete and the lowest possible excitation energy is set by the trap frequency (white horizontal line). For decreasing contact interaction strengths we can observe the roton minimum emerging, until finally its gap closes and the supersolid appears. In the supersolid regime we can clearly see the low-energy out-of-phase Goldstone mode and the large gap to all the other modes above the trap frequency of 30 Hz. The colour scale is normalized to the mode with the highest response across the scattering lengths shown.



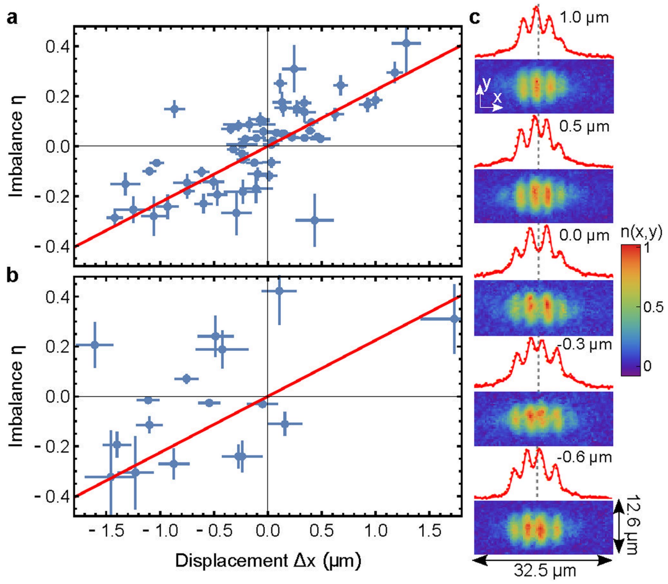
**Extended Data Fig. 3 | Phase pattern of the Goldstone mode.** Shown are the density cut through the three-droplet ground state (black) along the *x* axis and the phase pattern (red) corresponding to the low-energy Goldstone mode.



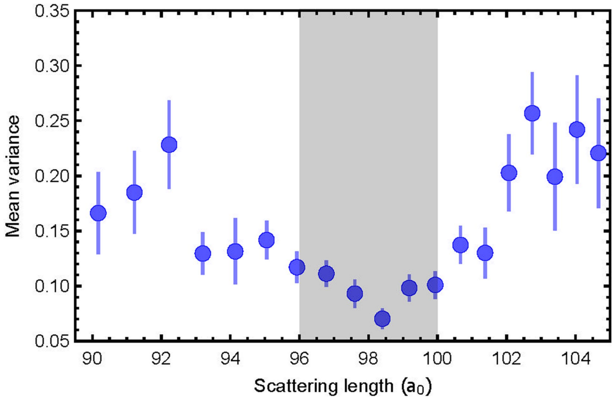
**Extended Data Fig. 4 | Large-amplitude dynamics of the out-of-phase Goldstone mode.** Starting from an array of three droplets, the state can change to a four-droplet state for large excitation amplitudes. From there it either oscillates back and forth between the two (**a**) or the excitation amplitude is so large that we find that the motion is no longer oscillatory, with the excitation instead travelling only in one direction (**b**). Similar to Fig. 2, $n(x)$ is the normalized line density along the $x$ axis.



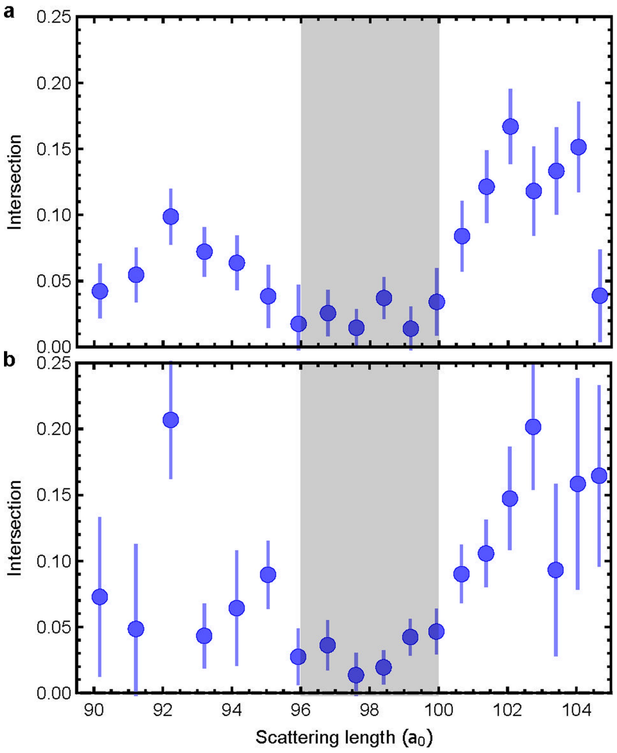
**Extended Data Fig. 5 | Simulated $\eta$–$\Delta x$ correlation of the four-droplet dynamic states.** Numerically predicted correlation between the imbalance $\eta$ and droplet displacement $\Delta x$ for the four-droplet states appearing at large excitation amplitudes of the low-energy Goldstone mode (blue points, with error bars indicating the uncertainty of the fit used to extract $\eta$ and $\Delta x$, as in Fig. 2). The red line is a linear fit.



**Extended Data Fig. 6 | Experimental correlation of the observed four-droplet states.** Similar to Fig. 3, we show the experimental correlation in the supersolid ($97.6a_0$, **a**) and isolated droplet ($91.2a_0$, **b**) regions, as well as example in situ images of four-droplet states, with $n(x,y)$ the normalized density (**c**). For the four-droplet states we observe a clear correlation of imbalance and displacement throughout the supersolid region.



**Extended Data Fig. 7 | Variance of the four-droplet data with respect to the theoretical correlation.** Similar to Fig. 4b, we find that the variance of the four-droplet data with respect to the theoretical correlation curve (Extended Data Fig. 6) is lowest in the supersolid region. As in Fig. 4a, the shaded area is the coherent region determined experimentally. The error bars indicate one standard error of the experimental data with respect to the theoretical correlation.



**Extended Data Fig. 8 | Theory-independent evaluation.** Intersection point of the fit with the displacement axis for the three-droplet states (**a**) and the four-droplet states (**b**) obtained from a linear fit to the $\eta$–$\Delta x$ correlation data across the part of the phase diagram we explored. An intersection point close to zero indicates the presence of a superfluid flow that can compensate fluctuations during the formation process. The error bars shown represent the standard error of the fitted intersections. As in Fig. 4a, the shaded area is the coherent region determined experimentally.

LETTER

https://doi.org/10.1038/s41586-019-1616-2

# A hexagonal planar transition-metal complex

Martí Garçon[1], Clare Bakewell[1], George A. Sackman[2,3], Andrew J. P. White[1], Richard I. Cooper[2], Alison J. Edwards[3] & Mark R. Crimmin[1]*

**Transition-metal complexes are widely used in the physical and biological sciences. They have essential roles in catalysis, synthesis, materials science, photophysics and bioinorganic chemistry. Our understanding of transition-metal complexes originates from Alfred Werner's realization that their three-dimensional shape influences their properties and reactivity[1], and the intrinsic link between shape and electronic structure is now firmly underpinned by molecular-orbital theory[2–5]. Despite more than a century of advances in this field, the geometries of transition-metal complexes remain limited to a few well-understood examples. The archetypal geometries of six-coordinate transition metals are octahedral and trigonal prismatic, and although deviations from ideal bond angles and bond lengths are frequent[6], alternative parent geometries are extremely rare[7]. The hexagonal planar coordination environment is known, but it is restricted to condensed metallic phases[8], the hexagonal pores of coordination polymers[9], or clusters that contain more than one transition metal in close proximity[10,11]. Such a geometry had been considered[12,13] for $[Ni(P^tBu)_6]$; however, an analysis of the molecular orbitals suggested that this complex is best described as a 16-electron species with a trigonal planar geometry[14]. Here we report the isolation and structural characterization of a simple coordination complex in which six ligands form bonds with a central transition metal in a hexagonal planar arrangement. The structure contains a central palladium atom surrounded by three hydride and three magnesium-based ligands. This finding has the potential to introduce additional design principles for transition-metal complexes, with implications for several scientific fields.**

A robust assignment of the coordination geometry of transition-metal complexes requires an understanding of both the position of the ligands and the nature of the chemical bonding. As an example, we consider six hydrogen atoms around a central metal in a planar arrangement[15]. Two extreme cases can be defined: either a trigonal planar geometry, in which the electrons reside within three H–H bonds; or a hexagonal planar geometry, in which all of the H–H bonds are broken and six new M–H interactions are formed. The two representations are limits of a continuum of bonding scenarios (Fig. 1a). Although both models are theoretical, the hexagonal planar geometry is considered to be highly implausible owing to the combination of the high formal oxidation state of the transition metal and the strongly σ-donating hydride ligands in the equatorial plane. We proposed that a plausible approach to obtain the hexagonal planar geometry would be to combine an alternating array of σ-donating ($L_\sigma$) and σ-accepting ligands ($L_{\sigma*}$) around the metal. This ligand topology would be expected to reduce the formal oxidation state of the metal[16–18] and favour the equatorial arrangement due to weak residual $L_\sigma$---$L_{\sigma*}$ interactions. Previous studies[19–22] have demonstrated that magnesium and zinc hydrides can coordinate to transition metals, and in certain cases the metal–hydride bond can break, leading to the formation of pairs of σ-donating and σ-accepting ligands.

The hexagonal planar complexes **1a** and **1b** were prepared by combining a suitable palladium precursor with a magnesium reagent that is stabilized by a β-diketiminate ligand (Fig. 1b; see Methods for synthetic details). Single-crystal X-ray diffraction studies were conducted (Fig. 1c), and the resulting data were of sufficient quality that electron density consistent with the hydride ligands could be located from the difference density map and their positions modelled by density functional theory (DFT) calculations[23]. Complexes **1a** and **1b** both possess a hexagonal planar geometry at palladium, with six ligands forming an equatorial plane in a near-perfect hexagonal arrangement. The Mg–Pd–H bond angles range between 54(2) and 67(2)°, with an average value of 60(2)° (the numbers in brackets correspond to one standard deviation). The sum of the angles around palladium is 360° for both **1a** and **1b**, and the largest deviation of the ligands away from the hexagonal plane is around 10°. To the best of our knowledge, crystallographically characterized Pd–Mg bonds are without precedent. Those in **1a** range from 2.550(1) Å to 2.567(1) Å, whereas those in **1b** are shorter at 2.485(1) Å to 2.497(1) Å; both are well within the sum of the single-bond covalent radii (Pauling[24], 2.64 Å; Pyykkö[25], 2.59 Å). The Pd–H bond lengths of **1a** and **1b** are short (1.57(4)–1.76(4) Å), whereas the Mg---H distances (2.08(5)–2.43(4) Å) are beyond the bonding limit of approximately 2.0 Å that has been established for

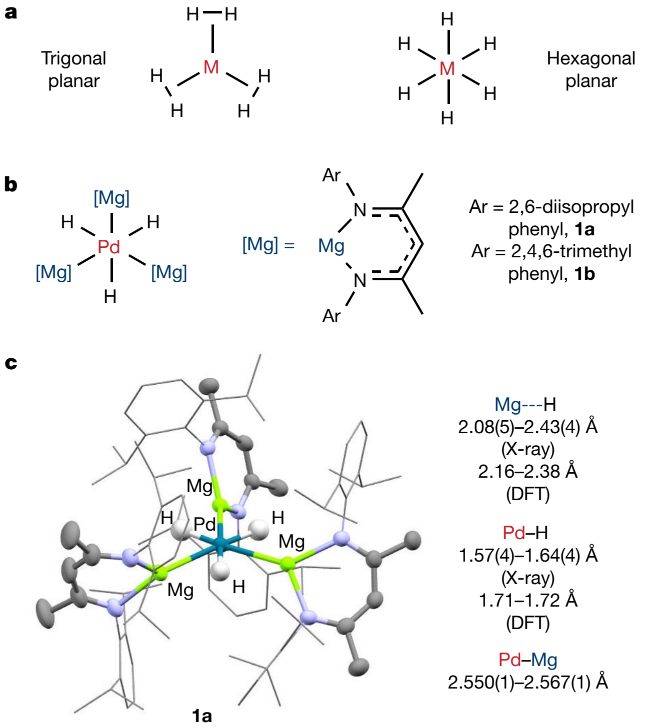


**Fig. 1** | **Preparation of hexagonal planar complexes. a**, Theoretical valence isomers of hypothetical $[MH_6]$ complexes showing two extreme bonding situations: left, a trigonal planar geometry; right, a hexagonal planar geometry. **b**, A line-drawing of the hexagonal planar complexes **1a** and **1b**, and a definition of the magnesium fragment [Mg]. **c**, Single-crystal X-ray derived model of **1a**, annotated with selected experimental and calculated bond lengths.

[1]Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, London, UK. [2]Chemical Crystallography, Chemistry Research Laboratory, University of Oxford, Oxford, UK. [3]Australian Centre for Neutron Scattering, Australian Nuclear Science and Technology Organisation, Sydney, New South Wales, Australia. *e-mail: m.crimmin@imperial.ac.uk

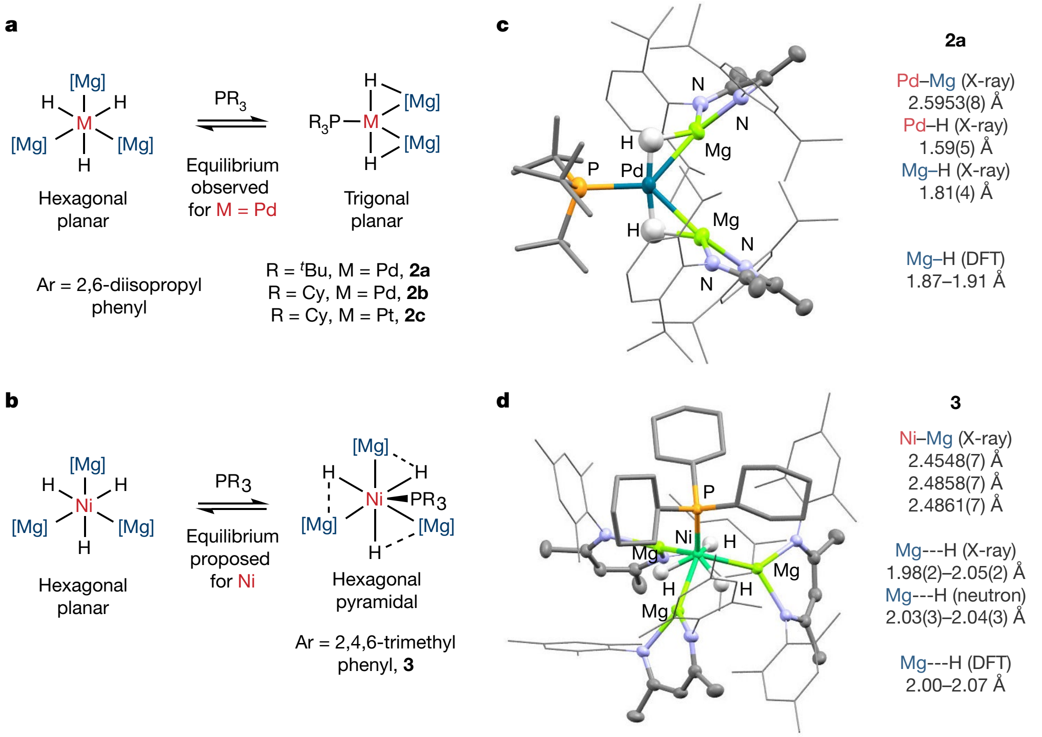


**Fig. 2 | Preparation of group-10 hydride complexes with magnesium ligands.** **a**, Line-drawings of the trigonal planar complexes **2a**–**c**, showing their relationship with the hexagonal planar analogues. **b**, A line-drawing of the hexagonal pyramidal complex **3**. **c**, **d**, Single-crystal X-ray-derived models of complexes **2a** and **3**, respectively, annotated with selected experimental (X-ray and neutron diffraction) and calculated bond lengths.

β-diketiminate-stabilized magnesium hydride complexes[26]. The closest Pd---H–C contacts in **1a** exceed 2.7 Å (Pd–H distance; Pd–C distances are greater than 3.6 Å), which rules out agostic interactions in the axial positions[27]. No Pd---H–C contacts are observed in the structure of **1b**.

Further synthetic experiments were carried out in which the transition-metal precursor was varied, enabling the isolation of complexes **2a–c** and **3** (Fig. 2; see Methods for synthetic details). This series of complexes provides points of comparison when examining the aforementioned hexagonal planar geometry. For example, in the solid state, **2a** possesses a distorted trigonal planar geometry. The palladium phosphine dihydride moiety of this complex is T-shaped with an H–Pd–H angle of 172(3)° and a P–Pd–H angle of 94(1)°. The Pd–Mg bond lengths in **2a** are approximately 0.05 Å longer than those in **1a**, while the Mg–H distances in **2a** are 0.3–0.6 Å shorter than the Mg---H separations in the hexagonal planar geometry and lie within the normal values expected for a bridging magnesium hydride. A similar structural motif has been observed in a cationic rhodium complex[21] and is found in the platinum analogue **2c**. By contrast, **3** contains an approximate hexagonal pyramidal geometry that is related to the hexagonal planar form through the association of an axial phosphine ligand. The Ni–Mg bond distances of **3** range from 2.455(1) Å to 2.486(1) Å and are similar to the sum of the single-bond covalent radii (Pauling radius[24], 2.51 Å; Pyykkö radius[25], 2.49 Å). For comparison, the Ni---Mg distances in the recently reported[28] [$Ni_5Mg$] cluster are longer and range from 2.562(10) Å to 2.947(13) Å. The hydride positions in **3** were unambiguously confirmed by a neutron diffraction study. The Mg---H distances of **3** (X-ray, 1.98(2)–2.05(2) Å; neutron, 2.03(3)–2.04(3) Å) are, at their longest, still 0.1–0.3 Å shorter than those observed in the hexagonal planar geometry.

The structures observed in the solid state persist in solution, as evidenced by multinuclear nuclear magnetic resonance (NMR) spectroscopy. In deuterated tetrahydrofuran (THF-$d_8$) solution at 273 K, diagnostic resonances of **3** were observed in the $^1$H and $^{31}$P{$^1$H} NMR spectra at chemical shifts of –7.27 ppm and 53.6 ppm, respectively. In $C_6D_6$ at 298 K, **1a** shows a single hydride resonance at –1.43 ppm, whereas **2b** is characterized by $^1$H and $^{31}$P resonances at –2.98 ppm and 37.1 ppm, respectively. At 283 K, the broad hydride peak of **2b** resolves into a doublet with a coupling constant, $^2J_{\mathrm{P-H}}$, of 9.8 Hz. This coupling constant lies within the 0–20 Hz range that has been established for square planar complexes containing phosphine and hydride ligands in a *cis* geometry[29]. The platinum analogue **2c** has a similar $^2J_{\mathrm{P-H}}$ of 8 Hz, with a $^1J_{\mathrm{Pt-H}}$ of 833 Hz and a $^1J_{\mathrm{Pt-P}}$ of 1,594 Hz. For comparison, in *trans*-[$Pt(H)_2(PCy_3)_2$] (Cy, cyclohexyl), the coupling constants[30] are $^2J_{\mathrm{P-H}} = 18.8$ Hz and $^1J_{\mathrm{Pt-H}} = 790$ Hz. The NMR data are consistent with the *trans* influence of the hydride ligands in **2b** and **2c** being similar to those found in related square planar complexes. Variable-temperature NMR experiments and titrations show that ligand exchange in this series of complexes is facile: **2b** exists in equilibrium with both **1a** and [$Pd(PCy_3)_2$], whereas phosphine dissociation was observed from **3**.

To better understand the hexagonal planar geometry, a series of calculations were undertaken. DFT calculations show that the Pd–Mg interactions in complexes **1a** and **2b** are predominantly ionic in nature. The magnesium atoms bear a substantial positive charge, while negative charge is localized primarily on the hydride ligands and, to a lesser extent, on the palladium centre. Wiberg bond indices (WBIs) provide an indication of the bond order and their evaluation enables us to assess the size of the covalent contribution to the bonds. The WBIs of both the Pd–Mg and the Mg---H interactions are higher for complex **2b** than for **1a**, indicating a stronger covalent interaction in **2b** than in **1a** (Fig. 3a). Quantum theory of atoms in molecules (QTAIM) calculations revealed a similar bonding picture. Whereas curved bond paths and associated bond-critical points were found between the magnesium and hydrogen atoms in **2b**, these features were all but absent in **1a**. Furthermore, **1a** displayed defined bond-critical points between the palladium and magnesium atoms that were not present for **2b** (Supplementary Fig. 21). In combination, these data enable the ligand–ligand interactions in the hexagonal planar complex to be quantified with some reliability. The calculations show that **1a** has the weakest Mg---H interactions of the complexes in the series.

Further insight into the bonding in **1a** can be obtained by considering two simple model complexes. The extreme trigonal planar and hexagonal planar geometries are conveniently described by the

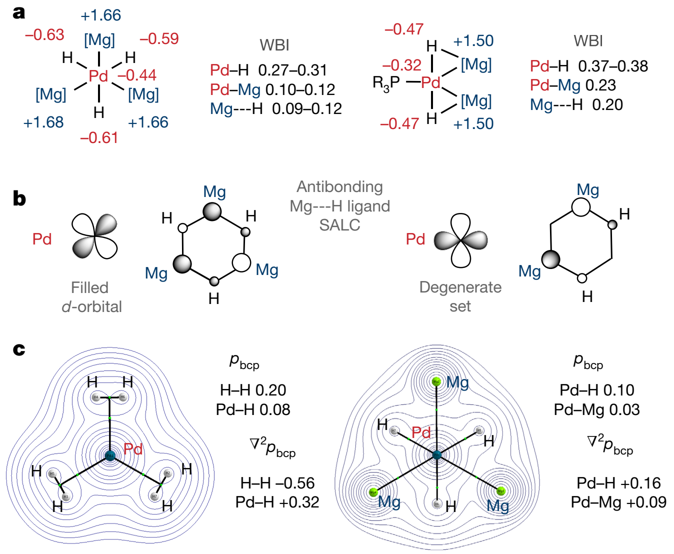


**Fig. 3 | Analysis of the chemical bonding in the hexagonal planar geometry.** **a**, Line-drawings of **1a** and **2b** showing the calculated natural population analysis charges on the key atoms; these charges show the accumulation of negative charge on palladium and hydrogen, and the accumulation of positive charge on magnesium. The WBIs, which indicate the strength of the covalent interaction, are shown on the right of the drawings. **b**, Key bonding interactions in the model complex $[Pd(H)_3(Mg)_3]^{3+}$, showing donor–acceptor interactions between palladium and magnesium or hydrogen atoms. **c**, Molecular graphs from QTAIM calculations on $[Pd(\eta^2\text{-}H_2)_3]$ (left) and $[Pd(H)_3(Mg)_3]^{3+}$ (right). $\rho_{bcp}$ indicates the electron density and $\nabla^2\rho_{bcp}$ is the Laplacian of the electron density at the bond-critical point.

16-electron complexes $[Pd(\eta^2\text{-}H_2)_3]$ and $[Pd(H)_3(Mg)_3]^{3+}$, respectively. These complexes can be assigned a formal Pd(0) oxidation state and $d^{10}$ electron count. Both can be assigned to the $D_{3h}$ point group and can be described by molecular orbital diagrams (Supplementary Figs. 26, 27). Inspection of the Kohn–Sham orbitals of the model complexes provides a qualitative description of the bonding. The key interactions that give rise to the hexagonal planar geometry of $[Pd(H)_3(Mg)_3]^{3+}$ are a doubly degenerate set of multi-centre two-electron bonds. In simple terms, these can be considered as donor–acceptor interactions between the filled $4d_{xy}$ and $4d_{x^2-y^2}$ orbitals and the corresponding empty ligand SALCs (symmetry adapted linear combinations) that are formed from the σ-acceptor ligands (Fig. 3b). The most relevant molecular orbitals in $[Pd(\eta^2\text{-}H_2)_3]$ are dominated by H–H bonding interactions that are constructed from the H 1*s* orbitals. QTAIM calculations returned a near-identical breakdown of the electronic structure (Fig. 3c). For $[Pd(\eta^2\text{-}H_2)_3]$, bond paths are found between pairs of hydrogen atoms in the ligand sphere. The electron density ($\rho_{bcp}$) values show that the H–H interactions are stronger than the Pd---H interactions, and the Laplacian of the electron density ($\nabla^2\rho_{bcp}$) shows a build-up of charge between the hydrogen atom pairs that is consistent with expectations for an H–H bond. For $[Pd(H)_3(Mg)_3]^{3+}$, bond-critical points are not found between the ligands but are present between palladium and magnesium and between palladium and hydrogen; the bond-critical paths radiate out in a hexagonal arrangement from the central metal atom. Scans of Mg---H bond lengths can be used to examine the difference between the hexagonal planar and trigonal planar geometries of $[Pd(H)_3(Mg)_3]^{3+}$. The potential energy surface that connects these two geometries is almost flat, which indicates that compression of the Mg---H interactions requires only a small amount of energy (Supplementary Fig. 23). Nevertheless, the hexagonal planar geometry is the global minimum on this surface for a series of computational methods.

In summary, the data we report here substantiate that complexes **1a** and **1b** are best described as 16-electron, Pd(0) complexes that have a hexagonal planar geometry. These complexes contain Pd–Mg bonds that are, to our knowledge, unprecedented. The valence electrons reside almost entirely in the metal–ligand bonds, with only weak residual interactions remaining between the hydride ($L_\sigma$) and magnesium ($L_{\sigma*}$) ligands. Minor perturbations of the structure to either include an axial phosphine ligand, to form **3**, or exchange an equatorial ligand, to form trigonal planar **2a**–**c**, lead to a measurable contraction of the Mg---H distances and a strengthening of the ligand---ligand interactions. The hexagonal planar geometry was originally predicted by Alfred Werner in seminal work determining the three-dimensional structures of transition-metal complexes through geometric isomerism[1]. The experimental realization of this coordination geometry fulfils a century-old hypothesis and has the potential to open up new and unconsidered possibilities across the physical and biological sciences.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1616-2.

Received: 7 June 2019; Accepted: 13 August 2019;

1. Werner, A. On the constitution and configuration of higher-order compounds. Nobel Lecture, 11 December 1913. *The Nobel Prize* https://www.nobelprize.org/prizes/chemistry/1913/werner/lecture (1913).
2. Hoffmann, R., Beier, B. F., Muetterties, E. L. & Rossi, A. R. Seven-coordination. A molecular orbital exploration of structure, stereochemistry, and reaction dynamics. *Inorg. Chem.* **16**, 511–522 (1977).
3. Rossi, A. R. & Hoffmann, R. Transition metal pentacoordination. *Inorg. Chem.* **14**, 365–374 (1975).
4. Gray, H. B. & Ballhausen, C. J. A molecular orbital theory for square planar metal complexes. *J. Am. Chem. Soc.* **85**, 260–265 (1963).
5. Gray, H. B. Molecular orbital theory for transition metal complexes. *J. Chem. Educ.* **41**, 2–12 (1964).
6. Alvarez, S. Distortion pathways of transition metal coordination polyhedra induced by chelating topology. *Chem. Rev.* **115**, 13447–13483 (2015).
7. García-Monforte, M. A., Baya, M., Falvello, L. R., Martín, A. & Menjón, B. An organotransition-metal complex with pentagonal-pyramidal structure. *Angew. Chem.* **51**, 8046–8049 (2012).
8. Yang, L.-M. et al. Two-dimensional $Cu_2Si$ monolayer with planar hexacoordinate copper and silicon bonding. *J. Am. Chem. Soc.* **137**, 2757–2762 (2015).
9. Niu, Z., Ma, J.-G., Shi, W. & Cheng, P. Water molecule-driven reversible single-crystal to single-crystal transformation of a multi-metallic coordination polymer with controllable metal ion movement. *Chem. Commun.* **50**, 1839–1841 (2014).
10. Tanabe, M. et al. Tetrapalladium complex with bridging germylene ligands. Structural change of the planar $Pd_4Ge_3$ core. *J. Am. Chem. Soc.* **133**, 18598–18601 (2011).
11. Yamada, T., Mawatari, A., Tanabe, M., Osakada, K. & Tanase, T. Planar tetranuclear and dumbbell-shaped octanuclear palladium complexes with bridging silylene ligands. *Angew. Chem.* **48**, 568–571 (2009).
12. Ahlrichs, R., Fenske, D., Oesen, H. & Schneider, U. Synthesis and structure of $[Ni(PtBu)_6]$ and $[Ni_5(PtBu)_6(CO)_5]$ and calculations on the electronic structure of $[Ni(PtBu)_6]$ and $(PR)_6$, R = *t*Bu,Me. *Angew. Chem.* **31**, 323–326 (1992).
13. Hey-Hawkins, E., Pink, M., Oesen, H. & Fenske, D. Synthesen und charakterisierung von $[Ni(tBuAs)_6]$ und $[Pd(tBuAs)_6]$. *Z. Anorg. Allg. Chem.* **622**, 689–691 (1996).
14. Tang, H., Hoffman, D. M., Albright, T. A., Deng, H. & Hoffmann, R. $[Ni(PtBu)_6]$ and $[Ni(SiH_2)_6]$ are isolobal, related to $[In\{Mn(CO)_4\}_5]^{2-}$, and have 16-electron counts. *Angew. Chem.* **32**, 1616–1618 (1993).
15. Kubas, G. J. Dihydrogen complexes as prototypes for the coordination chemistry of saturated molecules. *Proc. Natl Acad. Sci. USA* **104**, 6901–6907 (2007).
16. Crabtree, R. H. A new oxidation state for Pd? *Science* **295**, 288–289 (2002).
17. Chen, W., Shimada, S. & Tanaka, M. Synthesis and structure of formally hexavalent palladium complexes. *Science* **295**, 308–310 (2002).
18. Aullón, G., Lledós, A. & Alvarez, S. Hexakis(silyl)palladium(VI) or palladium(II) with $\eta^2$-disilane ligands? *Angew. Chem.* **41**, 1956–1959 (2002).
19. Butler, M. J. & Crimmin, M. R. Magnesium, zinc, aluminium and gallium hydride complexes of the transition metals. *Chem. Commun.* **53**, 1348–1365 (2017).
20. Abdalla, J. A. B. et al. Coordination and activation of Al–H and Ga–H bonds. *Chem. Eur. J.* **20**, 17624–17634 (2014).
21. Abdalla, J. A. B. et al. Structural snapshots of concerted double E–H bond activation at a transition metal centre. *Nat. Chem.* **9**, 1256–1262 (2017).
22. Riddlestone, I. M. et al. Activation of $H_2$ over the Ru–Zn bond in the transition metal–Lewis acid heterobimetallic species $[Ru(IPr)_2(CO)ZnEt]^+$. *J. Am. Chem. Soc.* **138**, 11081–11084 (2016).

23. Sharninghausen, L. S. et al. The neutron diffraction structure of $[Ir_4(IMe)_8H_{10}]^{2+}$ polyhydride cluster: testing the computational hydride positional assignments. *J. Organomet. Chem.* **849–850**, 17–21 (2017).
24. Pauling, L. Atomic radii and interatomic distances in metals. *J. Am. Chem. Soc.* **69**, 542–553 (1947).
25. Pyykkö, P. & Atsumi, M. Molecular single-bond covalent radii for elements 1–118. *Chem. Eur. J.* **15**, 186–197 (2009).
26. Mukherjee, D. & Okuda, J. Molecular magnesium hydrides. *Angew. Chem.* **57**, 1458–1473 (2018).
27. Brookhart, M., Green, M. L. H. & Parkin, G. Agostic interactions in transition metal compounds. *Proc. Natl Acad. Sci. USA* **104**, 6908–6914 (2007).
28. Shoshani, M. M., Liu, J. & Johnson, S. A. Mechanistic insight into H/D exchange by a pentanuclear Ni–H cluster and synthesis and characterization of structural analogues of potential intermediates. *Organometallics* **37**, 116–126 (2018).
29. Kudo, K., Hidai, M. & Uchida, Y. Dihydride complexes of platinum(II) and palladium(II). *J. Organomet. Chem.* **56**, 413–418 (1973).
30. Leviston, P. G. & Wallbridge, M. G. H. The preparation of some bulky dihalo-, halohydrido- and dihydro- phosphineplatinum(II) compounds. *J. Organomet. Chem.* **110**, 271–279 (1976).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

**General procedures.** Unless otherwise specified, all manipulations were carried out using standard Schlenk and glovebox techniques, under an inert atmosphere (nitrogen or argon). A MBRAUN Labmaster glovebox was used, operating at concentrations of $H_2O$ and $O_2$ of less than 0.1 ppm. Anhydrous solvents were obtained from a Grubbs-type solvent-purification system and stored over activated 3 Å molecular sieves in an inert atmosphere. Preparation of starting materials is detailed in Supplementary Information. All other reagents were obtained from commercial suppliers (Sigma-Aldrich, Alfa Aesar, Fluorochem) and used without further purification.

**Synthesis of 1a.** *Method 1.* [Pd(Me)$_2$($\kappa^2$-TMEDA)] (40 mg, 0.16 mmol, 1.0 equiv.) and [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ (Ar = 2,6-diisopropylphenyl, 189 mg, 0.21 mmol, 1.35 equiv.) were dissolved in toluene (10 ml). The reaction mixture immediately turned black and was stirred at 25 °C overnight. The volatiles were removed under vacuum and the residue washed with a small amount of cold *n*-hexane (1 ml). The solid was recrystallized from an approximately 1:1 mixture of toluene: *n*-hexane (2 ml) at −35 °C and the crystals washed successively with small amounts of cold *n*-hexane (3 × 0.5 ml), to afford the product **1a** as a dark solid (50 mg, 0.035 mmol, 22% yield). Reliable isolation of high-purity **1a** using this method is problematic, so an alternative method was developed.

*Method 2.* In a glovebox, [Pd($\eta^5$-Cp)($\eta^3$-cinnamyl)] (100 mg, 0.35 mmol, 1.0 equiv.) and [Al(H)$_2${(ArNCMe)$_2$CH}] (Ar = 2,6-diisopropylphenyl, 325 mg, 0.73 mmol, 2.1 equiv.) were dissolved in benzene (5 ml) to form a dark-brown solution, which was left to stand without stirring at 25 °C for 72 h. A red solid slowly precipitated. The mother liquor was decanted, and the solid was quickly washed with benzene (2 × 2 ml) and dried under vacuum to yield [PdAl(H)$_2${(ArNCMe)$_2$CH}]$_2$ as red crystals (95 mg, 0.086 mmol, 50% yield). This product was used in the next step without further purification. [PdAl(H)$_2${(ArNCMe)$_2$CH}]$_2$ (50 mg, 0.045 mmol, 1 equiv.) and [Mg($\mu$-H){$\kappa^2$-(ArNCMe)$_2$CH}]$_2$ (Ar = 2,6-diisopropylphenyl, 120 mg, 0.135 mmol, 3 equiv.) were suspended in benzene (5 ml) and the mixture was stirred at 50 °C for 72 h, during which time the suspension became a yellow-orange solution. The solvent was removed under vacuum and the residue washed with *n*-hexane (3 × 0.5 ml). The pale-brown solid was then dissolved in a 1:1 toluene:*n*-hexane mixture (1 ml) and filtered through a glass fibre to remove small amounts of a black solid, which was assumed to be Pd(0). The solution was stored at −35 °C and the product crystallized as colourless tablets. Complex **1a** was isolated as an off-white solid (33 mg, 0.023 mmol, 26% yield). $^1$H NMR (400 MHz, $C_6D_6$, 298 K): $\delta$ (ppm): −1.43 (s, 3H, Pd*H*$_3$), 0.96 (d, $^3J_{H-H}$ = 6.9 Hz, 36H, CH*Me*$_2$), 1.19 (d, $^3J_{H-H}$ = 6.7 Hz, 36H, CH*Me*$_2$), 1.53 (s, 18H, *Me*), 3.13 (sept, $^3J_{H-H}$ = 6.8 Hz, 12H, C*H*Me$_2$), 4.85 (s, 3H, β-C*H*), 6.85–7.40 (series of overlapping m, 18H, *Ar*). $T_1$ relaxation time (Pd*H*$_3$ signal, 298K): 0.83 s. $^{13}$C{$^1$H} NMR (100 MHz, $C_6H_6$): $\delta$ (ppm): 23.7 (12 × CH$_3$), 24.2 (6 × CH$_3$), 25.2 (12 × CH$_3$), 27.9 (12 × CH), 95.7 (3 × CH), 123.4 (12 × CH), 124.8 (6 × CH), 142.0 (12 × C), 145.4 (6 × C), 168.7 (6 × C). Attenuated total reflectance infrared spectroscopy (ATR IR; cm$^{-1}$): 3,056, 2,959, 2,926, 2,866, 1,618, 1,547, 1,431, 1,405, 1,364, 1,312, 1,252, 1,174, 1,100, 1,021, 757. Anal. Calc. ($C_{87}H_{126}Mg_3N_6Pd$): C, 72.80; H, 8.85; N, 5.86. Found: C, 72.66; H, 8.97; N, 5.74. Crystal data for **1a**: $C_{87}H_{126}Mg_3N_6Pd\cdot0.35(C_6H_{14})$, $M$ = 1,465.42, triclinic, *P*-1 (no. 2), $a$ = 13.2847(5), $b$ = 13.4347(5), $c$ = 24.7995(10) Å, $\alpha$ = 87.949(3), $\beta$ = 86.394(3), $\gamma$ = 76.189(3)°, $V$ = 4,288.7(3) Å$^3$, $Z$ = 2, $D_c$ = 1.135 g cm$^{-3}$, $\mu$(Cu-K$\alpha$) = 2.293 mm$^{-1}$, $T$ = 173 K, colourless tablets, Agilent Xcalibur PX Ultra A diffractometer; 16,321 independent measured reflections ($R_{int}$ = 0.0560), $F^2$ refinement, $R_1$(obs) = 0.0463, $wR_2$(all) = 0.1085, 12,341 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{max}$ = 147°], 981 parameters. CCDC: 1909687.

**Isolation of 1b.** In a J. Young's NMR tube, [{(MesNCMe)$_2$CH}Mg]$_2$ (Mes = 2,4,6-trimethylphenyl, 30 mg, 0.04 mmol, 1.0 equiv.) was suspended in dry toluene (1 ml) and [Pd(PCy$_3$)$_2$] (2.8 mg, 4.2 × 10$^{-3}$ mmol, 0.10 equiv.) was added. The reaction mixture was left for 48 h at 25 °C. Over this time period the reaction mixture turned from turbid to a clear yellow solution. The volatiles were then removed under vacuum and the crude product was redissolved in *n*-hexane (1 ml) and left at −35 °C. After repeated attempts, a few X-ray quality crystals of **1b** were obtained. Attempts to obtain **1b** on a preparative scale failed. Increasing the palladium loading resulted in regeneration of [Pd(PCy$_3$)$_2$]. Crystal data for **1b**: $C_{69}H_{90}Mg_3N_6Pd$, $M$ = 1,182.79, monoclinic, $P2_1/c$ (no. 14), $a$ = 25.9596(3), $b$ = 12.45210(14), $c$ = 21.1354(3) Å, $\beta$ = 100.3614(12), $V$ = 6,720.63(14) Å$^3$, $Z$ = 4, $D_c$ = 1.169 g cm$^{-3}$, $\mu$(Cu-K$\alpha$) = 2.817 mm$^{-1}$, $T$ = 173 K, colourless tablets, Agilent Xcalibur PX Ultra A diffractometer; 12,913 independent measured reflections ($R_{int}$ = 0.0269), $F^2$ refinement, $R_1$(obs) = 0.0449, $wR_2$(all) = 0.1343, 10,387 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{max}$ = 148°], 748 parameters. CCDC: 1909688.

**Synthesis of 2a.** In a glovebox, [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ (Ar = 2,6-diisopropylphenyl, 26 mg, 0.030 mmol, 1.25 equiv.) was added to a solution of [Pd(Me)$_2$($\kappa^2$-TMEDA)] (TMEDA, tetramethylethylamine; 6 mg, 0.024 mmol, 1 equiv.) in dry benzene in a J. Young's NMR tube. A solution of P$^t$Bu$_3$ (5 mg, 0.023 mmol, 0.95 equiv.) in benzene was then added dropwise. Formation of **2a** as well as [Pd(P$^t$Bu$_3$)$_2$] was identified by $^{31}$P NMR spectroscopy. The conversion did not increase after 18 h at 25 °C. The solvent was removed under vacuum. In the glovebox, the crude product was suspended in dry *n*-hexane, the excess starting material was removed by filtration with a polytetrafluoroethylene 0.2 μm HPLC filter and the clear brown solution was left at –35 °C to afford crystals of **2a** suitable for X-ray diffraction. $^1$H NMR (500 MHz, $C_6H_6$, 283 K): $\delta$ (ppm): –4.49 (s, 2H, Pd*H*$_2$). $^{31}$P{$^1$H} NMR (202 MHz, $C_6H_6$, 283 K): $\delta$ (ppm): 86.6 (broad s). Crystal data for **2a**: $C_{70}H_{111}Mg_2N_4PPd\cdot2(C_6H_{14})$, $M$ = 1,366.95, monoclinic, *I*2/*a* (no. 15), $a$ = 15.52641(16), $b$ = 20.9811(2), $c$ = 24.8465(3) Å, β = 90.7595(9)°, $V$ = 8093.32(14) Å$^3$, $Z$ = 4 (half a molecule per asymmetric unit), $D_c$ = 1.122 g cm$^{-3}$, $\mu$(Cu-K$\alpha$) = 2.494 mm$^{-1}$, $T$ = 173 K, brown blocks, Agilent Xcalibur PX Ultra A diffractometer; 8,015 independent measured reflections ($R_{int}$ = 0.0357), $F^2$ refinement, $R_1$(obs) = 0.0435, $wR_2$(all) = 0.1220, 7,154 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{max}$ = 148°], 500 parameters. CCDC: 1589324. Solution stability: Complex **2a** is unstable in solution and forms an equilibrium mixture of **2a**, [Pd(P$^t$Bu$_3$)$_2$] and [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$.

**Synthesis of 2b.** *Method 1.* In a glovebox, [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ (Ar = 2,6-diisopropylphenyl, 50 mg, 0.057 mmol, 2.5 equiv.) and [Pd(PCy$_3$)$_2$] (15 mg, 0.022 mmol, 1 equiv.) were dissolved in dry benzene ($C_6H_6$) in a J. Young's NMR tube. The conversion did not increase after 5 h at 25 °C to give an approximately 1:6 mixture of [Pd(PCy$_3$)$_2$]:**2b**, as evidenced by $^{31}$P-NMR spectroscopy. The solvent was removed under vacuum. In the glovebox, the crude product was suspended in dry *n*-hexane (1 ml), unreacted [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ was removed by filtration with a polytetrafluoroethylene 0.2 μm HPLC filter and the clear yellow solution was left at –35 °C to afford crystals of **2b** suitable for X-ray diffraction.

*Method 2.* In a glovebox, [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ (Ar = 2,6-diisopropylphenyl, 47 mg, 0.053 mmol, 2.25 equiv.) was added to a solution of [Pd(Me)$_2$($\kappa^2$-TMEDA)] (6 mg, 0.024 mmol, 1 equiv.) in dry benzene in a J. Young's NMR tube. A solution of PCy$_3$ (7 mg, 0.024 mmol, 1 equiv.) in benzene was then added dropwise. Immediate formation at 25 °C of **2b** as well as [Pd(PCy$_3$)$_2$] was identified by $^{31}$P NMR spectroscopy. $^1$H NMR (500 MHz, $C_6H_6$, 298 K): $\delta$ (ppm): −2.98 (s, 2H, Pd*H*$_2$). Partial data only.$^1$H NMR (400 MHz, $C_7D_8$, 233 K) $\delta$ (ppm): –2.60 (br s, 2H, Pd*H*$_2$), 0.35 (d, $^3J_{H-H}$ = 6.4 Hz, 6H, CH*Me*$_2$), 0.76–2.21 (series of overlapping m, CH*Me*$_2$, *Me*, P*Cy*$_3$), 2.87 (sept, $^3J_{H-H}$ = 6.4 Hz, 2H, C*H*Me$_2$), 3.27 (sept, $^3J_{H-H}$ = 6.8 Hz, 2H, C*H*Me$_2$), 3.44 (sept, $^3J_{H-H}$ = 6.8 Hz, 2H, C*H*Me$_2$), 3.49 (sept, $^3J_{H-H}$ = 6.4 Hz, 2H, C*H*Me$_2$), 4.85 (s, 2H, β-C*H*), 6.95–7.20 (series of overlapping m, *Ar*). $^{31}$P{$^1$H} NMR (202 MHz, $C_6H_6$): $\delta$ (ppm): 37.08 (s). $T_1$ relaxation time (Pd*H*$_2$ signal, 298K): 0.60 s. Crystal data for **2b**: $C_{76}H_{117}Mg_2N_4PPd\cdot C_6H_{14}$, $M$ = 1,358.89, triclinic, *P*-1 (no. 2), $a$ = 14.8889(15), $b$ = 16.1336(10), $c$ = 17.0165(11) Å, $\alpha$ = 96.422(5), $\beta$ = 97.216(7), $\gamma$ = 97.942(7)°, $V$ = 3,981.4(6) Å$^3$, $Z$ = 2, $D_c$ = 1.134 g cm$^{-3}$, $\mu$(Cu-K$\alpha$) = 2.534 mm$^{-1}$, $T$ = 173 K, pale-yellow plates, Agilent Xcalibur PX Ultra A diffractometer; 15,070 independent measured reflections ($R_{int}$ = 0.0850), $F^2$ refinement, $R_1$(obs) = 0.0685, $wR_2$(all) = 0.1857, 10,575 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{max}$ = 148°], 833 parameters. CCDC: 1589323. Solution stability: Complex **2b** forms a complex equilibrium mixture in solution with not only [Pd(PCy$_3$)$_2$] and [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ but also **1a** in hydrocarbon solutions at 25 °C.

**Synthesis of 2c.** In a glovebox, [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ (58.4 mg, 0.066 mmol, 1.0 equiv.) and [Pt(PCy$_3$)$_2$] (50 mg, 0.066 mmol, 1.0 equiv.) were dissolved in dry benzene (3 ml) in a small ampoule. The resulting yellow solution was stirred at 25 °C for 3 h. The solvent was then removed under vacuum. The crude product was dissolved in a small amount of dry *n*-hexane (0.5 ml), the excess [Mg($\mu$-H){(ArNCMe)$_2$CH}]$_2$ was removed by filtration through a glass fibre and the clear yellow solution was stored at −35 °C to afford the desired product **2c** as a pale-yellow microcrystalline solid (85 mg, 0.062 mmol, 95% yield). $^1$H NMR (500 MHz, $C_6D_6$, 298 K): $\delta$ (ppm): −6.25 (d, $^2J_{H-P}$ = 8.5 Hz and $^1J_{H-Pt}$ = 833 Hz (satellites), 2H, Pt*H*$_2$), 0.48 (br s, 3H), 1.03–1.95 (series of overlapping m, CH*Me*$_2$, *Me*, P*Cy*$_3$), 2.77 (br sept, $^3J_{H-H}$ = 6.4 Hz, 2H, C*H*Me$_2$), 3.39 (br sept, $^3J_{H-H}$ = 6.8 Hz, 2H, C*H*Me$_2$), 3.47 (br sept, $^3J_{H-H}$ = 6.8 Hz, 2H, C*H*Me$_2$), 3.56 (br sept, $^3J_{H-H}$ = 6.4 Hz, 2H, C*H*Me$_2$), 4.92 (s, 2H, β-C*H*), 6.91–7.30 (series of overlapping m, *Ar*). $T_1$ relaxation time (Pt*H*$_2$ signal, 298 K): 0.87 s. $^{31}$P{$^1$H} NMR (202 MHz, $C_6H_6$): $\delta$ (ppm): 44.04 (s, $^1J_{P-Pt}$ = 1,600 Hz (satellites)). $^{195}$Pt{$^1$H} NMR (107.5 Hz, $C_6H_6$): $\delta$ (ppm): −5,619.1 (d, $^1J_{P-Pt}$ = 1,600 Hz). $^{13}$C{$^1$H} NMR (126 MHz, $C_6H_6$): $\delta$ (ppm): 23.7–29.2 (series of overlapping signals, CH*Me*$_2$, *Me*, P*Cy*$_3$, CHMe$_2$), 30.8 (br s, 6 × CH$_2$), 38.0 (d, $^1J_{C-P}$ = 12 Hz, 3 × CH), 97.8 (2 × CH), 123.2 (2 × CH), 123.3 (2 × CH), 124.0 (2 × CH), 124.1 (2 × CH), 125.3 (2 × CH), 125.8 (2 × CH), 142.2 (2 × C), 142.4 (2 × C), 143.1 (2 × C), 144.0 (2 × C), 146.3 (2 × C), 147.0 (2 × C), 168.3 (2 × C), 168.7 (2 × C). ATR IR (cm$^{-1}$): 3,058, 2,957, 2,924, 2,850, 1,661, 1,622, 1,550, 1,524, 1,460, 1,435, 1,408, 1,382, 1,362, 1,314, 1,254, 1,174, 1,100, 1,020, 793, 759. Anal. Calc. ($C_{76}H_{117}Mg_2N_4PtP$): C, 67.05; H, 8.66; N, 4.12. Found: C, 66.40; H, 8.77; N, 3.86. Crystal data for **2c**: $C_{76}H_{117}Mg_2N_4PPt\cdot1.35(C_6H_{14})$, $M$ = 1,477.74, triclinic, *P*-1 (no. 2), $a$ = 13.2111(3), $b$ = 15.1411(4), $c$ = 42.2615(12) Å, $\alpha$ = 83.607(2),

$\beta = 88.005(2)$, $\gamma = 78.713(2)°$, $V = 8{,}237.8(4)$ Å$^3$, $Z = 4$ (2 molecules per asymmetric unit), $D_c = 1.192$ g cm$^{-3}$, $\mu$(Cu-K$\alpha$) = 3.828 mm$^{-1}$, $T = 173$ K, colourless needles, Agilent Xcalibur PX Ultra A diffractometer; 31,579 independent measured reflections ($R_{\mathrm{int}} = 0.0426$), $F^2$ refinement, $R_1$(obs) = 0.0407, $wR_2$(all) = 0.0977, 23,332 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{\mathrm{max}} = 148°$], 1,630 parameters. CCDC: 1909689.

**Synthesis of 3.** In a glovebox, [{(MesNCMe)$_2$CH}Mg]$_2$ (100 mg, 0.14 mmol, 4.0 equiv.) and [Ni(PCy$_3$)$_2$]$_2$($\mu$-N$_2$) (44 mg, 0.035 mmol, 1.0 equiv.) were transferred to a J. Young's tap ampoule. Benzene (5 ml) was added and the reaction was heated at 80 °C for 16 h. The reaction mixture was allowed to cool to 25 °C and *n*-hexane (5 ml) was added. Yellow crystals of **3** formed after standing overnight at 25 °C. The crystals were isolated by filtration and washed with *n*-hexane to give **3** as a yellow solid (93 mg, 0.066 mmol, 95% yield). $^1$H NMR (400 MHz, THF-$d_8$, 273 K): $\delta$ (ppm): −7.27 (s, 3H, Mg–*H*–Ni), 0.06 (m, 3H), 0.42 (m, 3H), 0.52 (m, 3H), 0.75–0.94 (m, 12H), 1.01 (m, 3H), 1.36 (s, 9H), 1.41 (s, 9H), 1.54 (s, 9H), 1.54 (s, 9H), 1.57 (m, 6H), 1.95 (s, 9H), 2.07 (s, 9H), 2.10 (s, 9H), 2.23 (s, 9H), 2.46 (s, 9H), 5.09 (s, 3H), 6.60 (s, 3H), 6.65 (s, 3H), 6.80 (s, 3H), 7.07 (s, 3H). $^{31}$P{$^1$H} NMR (162 MHz, THF-$d_8$, 273 K): $\delta$ (ppm): 53.6 (s, *P*Cy$_3$). $^{13}$C{$^1$H} NMR (100 MHz, THF-$d_8$, 273 K): $\delta$ (ppm): 19.6 (*C*H$_3$), 19.8 (*C*H$_3$), 20.3 (*C*H$_3$), 20.7 (*C*H$_3$), 21.2 (*C*H$_3$), 22.0 (*C*H$_3$), 23.5 (*C*H$_3$), 24.4 (*C*H$_3$), 27.6, 28.1, 28.2, 28.5, 28.6, 30.2, 31.3, 31.4, 32.5, 35.1, 35.2, 98.0, 128.8 (*C*), 128.9 (*C*H), 130.2 (*C*H), 130.2 (*C*H), 130.8 (*C*H), 131.7 (*C*), 132.6 (*C*), 132.8 (*C*), 132.9 (*C*), 133.4 (*C*), 147.2 (*C*), 148.4 (*C*), 168.2 (*C*), 168.3 (*C*). Anal. Calc. ($C_{87}H_{123}Mg_3N_6NiP$): C, 73.82; H, 8.76; N, 5.94. Found: C, 73.67; H, 8.82; N, 5.59. Crystal data for **3**: X-ray: $C_{87}H_{123}Mg_3N_6NiP{\cdot}3(C_6H_6)$, $M = 1{,}649.84$, triclinic, *P*-1 (no. 2), $a = 14.2326(5)$, $b = 14.4809(4)$, $c = 24.0822(9)$ Å, $\alpha = 78.865(3)$, $\beta = 79.981(3)$, $\gamma = 84.955(2)°$, $V = 4{,}788.0(3)$ Å$^3$, $Z = 2$, $D_c = 1.144$ g cm$^{-3}$, $\mu$(Cu-K$\alpha$) = 1.004 mm$^{-1}$, $T = 173$ K, yellow blocks, Agilent Xcalibur PX Ultra A diffractometer; 18,329 independent measured reflections ($R_{\mathrm{int}} = 0.0393$), $F^2$ refinement, $R_1$(obs) = 0.0439, $wR_2$(all) = 0.1154, 13,714 independent observed absorption-corrected reflections [$|F_o| > 4\sigma(|F_o|)$, $2\theta_{\mathrm{max}} = 148°$], 1,084 parameters. CCDC: 1909690. Neutron diffraction data for **3**: See Supplementary Information for full details. Neutron Laue data were collected on the KOALA instrument at the Australian Nuclear Science and Technology Organisation (ANSTO). A total of 102,818 reflections with wavelengths between 0.85 Å and 1.70 Å covering the full sphere of reciprocal space to a maximum resolution of 0.98 Å were reduced to yield 8,847 independent reflections [L4R(int) = 0.10(7)]; 4,816 with $I > 3\sigma(I)$. Refinement of a structure model in CRYSTALS converged to $R_1 = 0.1117$, $wR_2 = 0.0985$ for $I > 3\sigma(I)$. Maximum and minimum residual difference densities: 1.16 and −1.50 fm Å$^{-3}$. CCDC: 1946045.

## Data availability

Crystallographic models are available as .cif files from the Cambridge Crystallographic Data Centre (CCDC, https://www.ccdc.cam.ac.uk); CCDC numbers 1589323–1589324, 1909687–1909690 and 1946045). Neutron diffraction images can be obtained from A.J.E. The derived structure factors have been deposited with the CCDC, with CCDC number 1946045. Data associated with DFT calculations (.xyz coordinate file) along with NMR spectroscopic data (.mnova files) are available from a public repository at https://doi.org/10.14469/hpc/5985. Full details of the syntheses are provided in the Supplementary Information.

**Acknowledgements** We thank the Royal Society and the European Research Council (FluoroFix: 677367) for funding, Johnson Matthey for the gift of PdCl$_2$, and ANSTO for allocation of neutron beam time on KOALA to proposal P6932. We thank Oxford Cryosystems Ltd and ANSTO for funding a Russell Studentship with the University of Oxford (GAS).

**Author contributions** M.G. and C.B. carried out the synthetic studies. M.G. conducted DFT, QTAIM and related calculations. A.J.P.W. collected, processed and refined single-crystal X-ray diffraction data. A.J.E., R.I.C. and G.A.S. undertook the single-crystal Laue neutron diffraction experiment, data reduction and refined the neutron diffraction structural model. M.R.C. managed the project. All authors contributed to the writing and editing of the manuscript.

**Competing interests** The authors declare no competing interests.

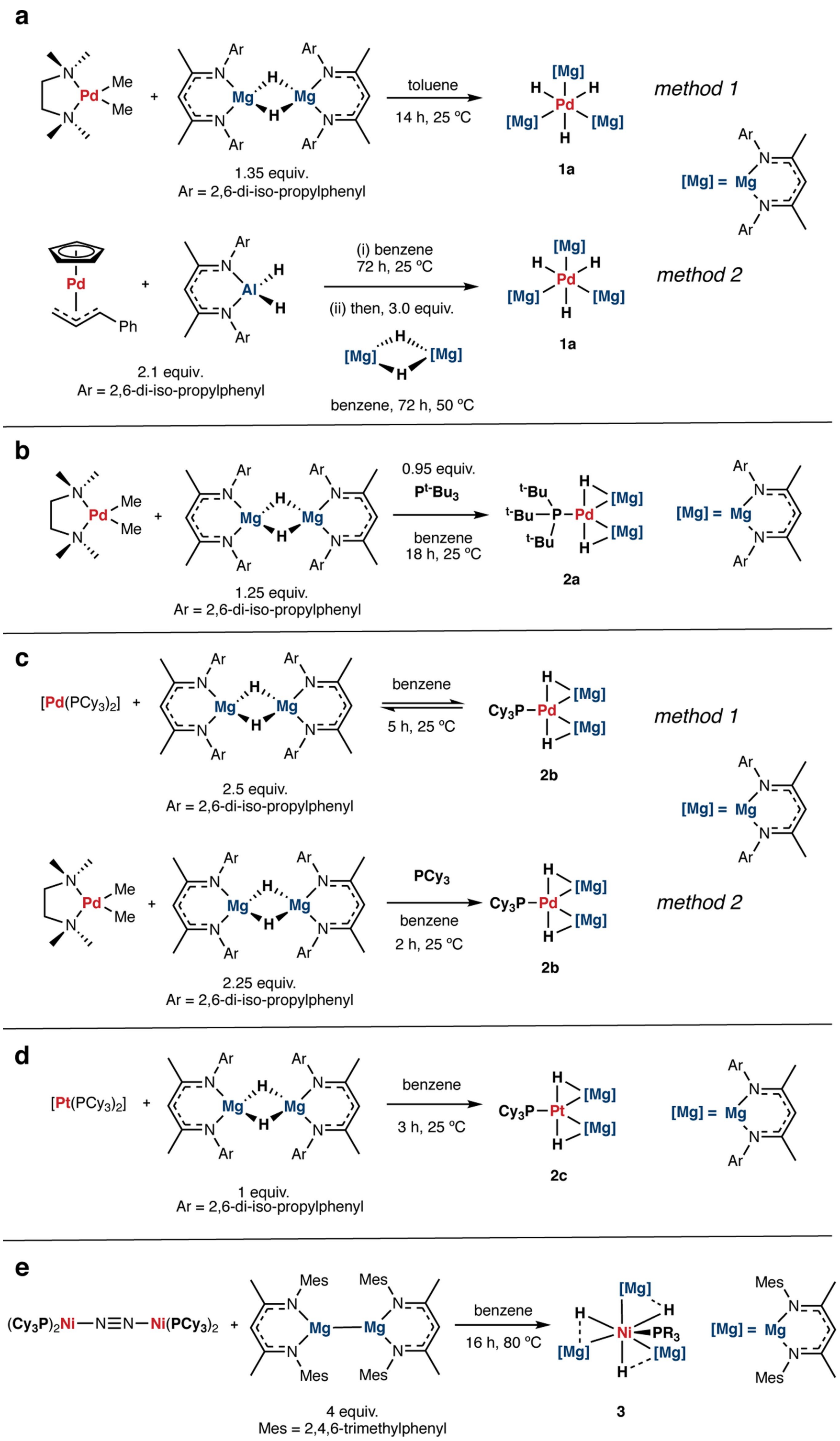
**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1616-2.

**Correspondence and requests for materials** should be addressed to M.R.C.

**Peer review information** *Nature* thanks Jean-François Halet and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints.



**Extended Data Fig. 1 | Synthesis of group-10 hydride complexes with magnesium ligands. a–e**, Synthetic schemes for the preparation of complexes **1a** (**a**), **2a** (**b**), **2b** (**c**), **2c** (**d**) and **3** (**e**); for **3**, the hydride ligands are derived from the C–H bonds of the benzene solvent.

LETTER

https://doi.org/10.1038/s41586-019-1645-x

# Crosslinking ionic oligomers as conformable precursors to calcium carbonate

Zhaoming Liu[1], Changyu Shao[1], Biao Jin[1], Zhisen Zhang[2], Yueqi Zhao[1], Xurong Xu[3] & Ruikang Tang[1,4]*

**Inorganic materials have essential roles in society, including in building construction, optical devices, mechanical engineering and as biomaterials[1–4]. However, the manufacture of inorganic materials is limited by classical crystallization[5], which often produces powders rather than monoliths with continuous structures. Several precursors that enable non-classical crystallization—such as prenucleation clusters[6–8], dense liquid droplets[9,10], polymer-induced liquid precursor phases[11–13] and nanoparticles[14]—have been proposed to improve the construction of inorganic materials, but the large-scale application of these precursors in monolith preparations is limited by availability and by practical considerations. Inspired by the processability of polymeric materials that can be manufactured by crosslinking monomers or oligomers[15], here we demonstrate the construction of continuously structured inorganic materials by crosslinking ionic oligomers. Using calcium carbonate as a model, we obtain a large quantity of its oligomers $(CaCO_3)_n$ with controllable molecular weights, in which triethylamine acts as a capping agent to stabilize the oligomers. The removal of triethylamine initiates crosslinking of the $(CaCO_3)_n$ oligomers, and thus the rapid construction of pure monolithic calcium carbonate and even single crystals with a continuous internal structure. The fluid-like behaviour of the oligomer precursor enables it to be readily processed or moulded into shapes, even for materials with structural complexity and variable morphologies. The material construction strategy that we introduce here arises from a fusion of classic inorganic and polymer chemistry, and uses the same crosslinking process for the manufacture the materials.**

Many materials are consolidated from their crystallized powders[16], but their resulting discontinuous internal structures render them brittle with a poor ability to resist fracture[17,18]. By contrast, polymeric materials are ubiquitous in modern society, due not only to their varied properties but also to their ease of fabrication[15,19]. The polymerization strategy is superior to crystallization because of its efficiency and controllability. In polymer chemistry, covalent bonds have an important role in ensuring the linkage of small units. Although a few covalent-bond-based inorganic materials (for example silicone and silica)[20,21] can be obtained as polymers, there is no general method for the preparation of such materials by crosslinking owing to the lack of investigation into ionic monomers or oligomers for this purpose. In the control of polymerization reactions, a capping agent is key[22]: capping can stabilize precursors, whereas de-capping can initiate polymerization. Analogously, we proposed that ionic oligomers could be stabilized

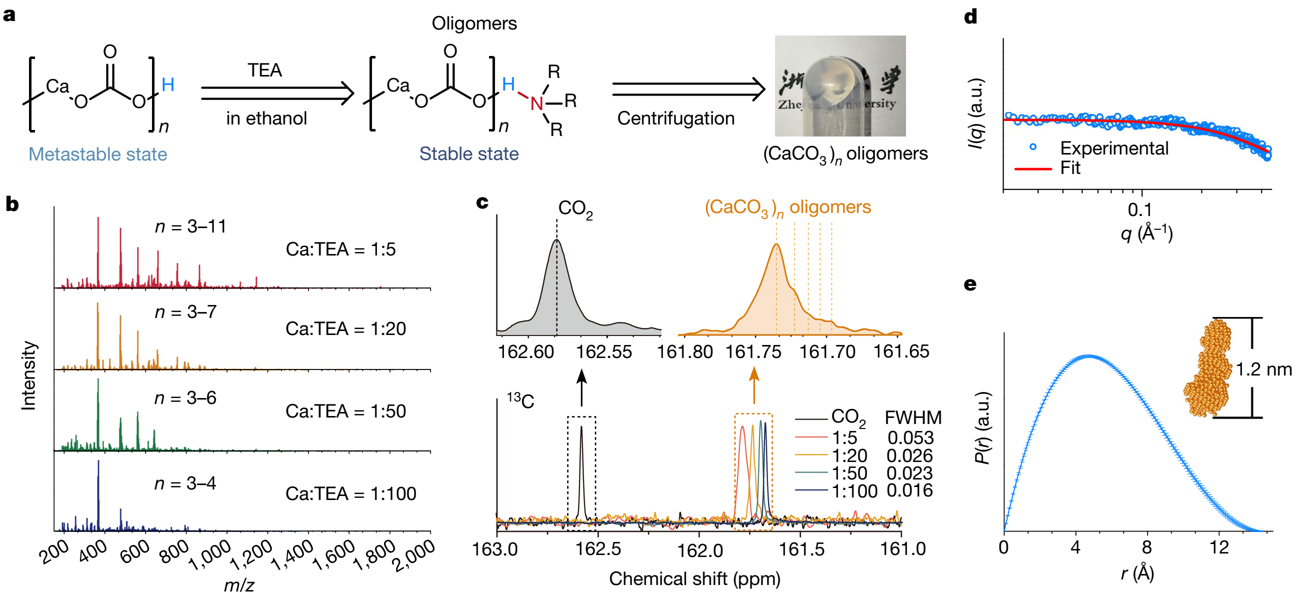


**Fig. 1 | Preparation and characterization of $(CaCO_3)_n$ oligomers.** **a**, Left, scheme of the capping strategy and reaction conditions for producing $(CaCO_3)_n$ oligomers; right, a photograph of gel-like $(CaCO_3)_n$ oligomers. **b**, Mass spectra of $(CaCO_3)_n$ oligomers with different Ca:TEA molar ratios. **c**, Liquid-state $^{13}C$ NMR spectra of $CO_2$ or the carbonates of $(CaCO_3)_n$ oligomers with different Ca:TEA molar ratios in ethanol. **d**, Scattering plots of $(CaCO_3)_n$ measured by SAXS. The red curve is the fitting result obtained using DAMMIF. *I*, scattering intensity; *q*, scattering vector. The error bar represents the standard deviation of twenty measurements. **e**, Pair–distance distribution function ($P(r)$) of the $(CaCO_3)_n$ oligomers. The inset shows the shape simulation of the oligomer. Error bars represent one standard deviation, $n = 20$.

---

[1]Department of Chemistry, Zhejiang University, Hangzhou, China. [2]Department of Physics, Research Institute for Biomimetics and Soft Matter, Jiujiang Research Institute, Xiamen University, Xiamen, China. [3]Qiushi Academy for Advanced Studies, Zhejiang University, Hangzhou, China. [4]State Key Laboratory of Silicon Materials, Zhejiang University, Hangzhou, China. *e-mail: rtang@zju.edu.cn

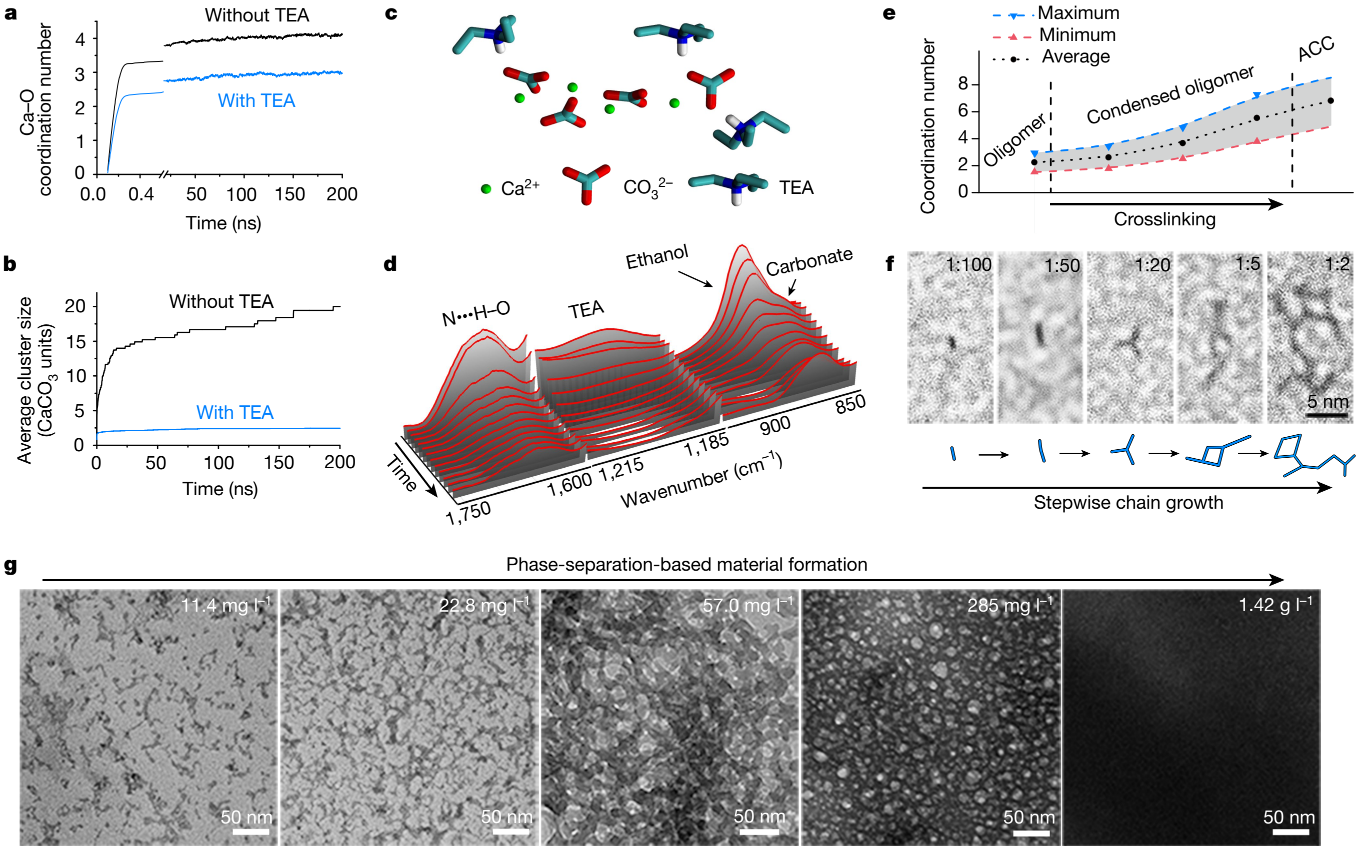


**Fig. 2 | Controllable crosslinking of $(CaCO_3)_n$ oligomers. a, b**, Molecular dynamics simulation of the evolution of the Ca–O (from carbonate) coordination number (**a**) and the average cluster size (**b**) from ions ($Ca^{2+}$ and $CO_3^{2-}$) in the absence (black) or presence (blue) of TEA. **c**, A typical simulated $CaCO_3$ cluster capped with TEA (an oligomer). **d**, In situ FTIR spectra during the drying of $(CaCO_3)_n$ oligomers. **e**, The change in the coordination number of Ca–O during crosslinking. Owing to the uncertainty in the exact density during measurements, the blue and red lines are shown to represent the maximum and minimum coordination number of Ca–O, respectively. The black line shows the average coordination number. **f**, High-resolution TEM images of $(CaCO_3)_n$ oligomers grown at different Ca:TEA ratios from 1:100 to 1:2. **g**, TEM images depicting the transformation of $(CaCO_3)_n$ oligomers to larger structures during condensation.

by an appropriate capping agent. Capping based on hydrogen bonding was thought to be suitable, because most inorganic complexes contain oxygen. For example, triethylamine (TEA) can form a hydrogen bond with a protonated carbonate through its tertiary amine group. More importantly, TEA is a small molecule that can be volatilized at room temperature, and it was expected that this could initiate an expected crosslinking reaction.

We chose calcium carbonate oligomers as the inorganic ionic unit. These oligomers can be generated by bubbling $CO_2$ into an ethanol solution containing calcium chloride dihydrate and TEA. Ethanol was used as the solvent because its low dielectric constant ($\varepsilon = 24.5$) favours the formation of a hydrogen bond between the nitrogen of TEA and the protonated carbonate[23,24] (Fig. 1a). The resulting oligomers ($(CaCO_3)_n$, in which $n$ represents the number of $Ca^{2+}$:$CO_3^{2-}$ units), were detected by electrospray ionization mass spectrometry (ESI-MS). The value of $n$ could be tuned by adjusting the molar ratio of calcium ions to TEA (Ca:TEA; Fig. 1b, Extended Data Fig. 1). For Ca:TEA = 1:5, peaks were observed at 368, 476, 564, 662, 756, 865, 1,065 and 1,143 $m/z$, implying $n$ values of 3–11, respectively (no peak for $n = 9$ was found). Upon increasing the concentration of TEA, the values of $n$ decreased to 3–7, 3–6 and finally 3–4 at Ca:TEA ratios of 1:20, 1:50 and 1:100, respectively. These results are consistent with an enhanced capping effect arising from increasing amounts of TEA.

Analysis by liquid-state nuclear magnetic resonance (NMR) spectroscopy (Fig. 1c) confirmed the formation of oligomers, in which the carbonates are in different chemical states. The chemical shift of $^{13}C$ from $CO_2$ in TEA-containing ethanol is 162.58 ppm, and the peak is sharp, whereas the value for carbonate in the resulting $(CaCO_3)_n$ (Ca:TEA = 1:20) is shifted to 161.73 ppm—owing to its binding to $Ca^{2+}$—with several shoulder peaks from the oligomers localized between 161.73 and 161.68 ppm. With increasing TEA concentration, the full width at half maximum (FWHM) of these NMR signals decreases from 0.053 (Ca:TEA = 1:5) to 0.016 (Ca:TEA = 1:100). This peak narrowing represents the decreased size of the oligomers (that is, a smaller value of $n$), which is consistent with the decreased $m/z$ ratio from mass spectrometry data. The sizes and shapes of $(CaCO_3)_n$ oligomers were examined at different concentrations using synchrotron small-angle X-ray scattering (SAXS, Fig. 1d, e, Extended Data Fig. 2a). We used oligomers formed at a Ca:TEA ratio of 1:100, as this ratio gave the narrowest distribution of products according to mass spectrometry experiments. After background subtraction, the scattering plots obtained at a concentration of 11.4 mg $l^{-1}$ showed no obvious inter-oligomer correlation (Fig. 1d), suggesting that the solution of $(CaCO_3)_n$ was monodisperse. Analysis using DAMMIF[25], a program that enables the shape of a substrate to be determined from SAXS data, showed that the oligomers were rod-like with a length of 1.2 nm (Fig. 1e); this is in agreement with the results of mass spectrometry, from which an $n$ value of 3–4 was obtained. The oligomers aggregate at increasing concentrations in ethanol; however, this process is reversible when they are capped by TEA, as dilution induces disaggregation back to individual oligomers (Extended Data Fig. 2). The gel-like oligomers can be concentrated from ethanol solution by high-speed centrifugation (27,500$g$) (Fig. 1a). Gravimetric analysis indicated that the gel was around 20 wt% inorganic $(CaCO_3)_n$ phase, and almost all of the residual mass was ethanol. Fourier transform infrared (FTIR) spectroscopy revealed the existence of $(CaCO_3)_n$, TEA, and an N···H–O hydrogen-bonding

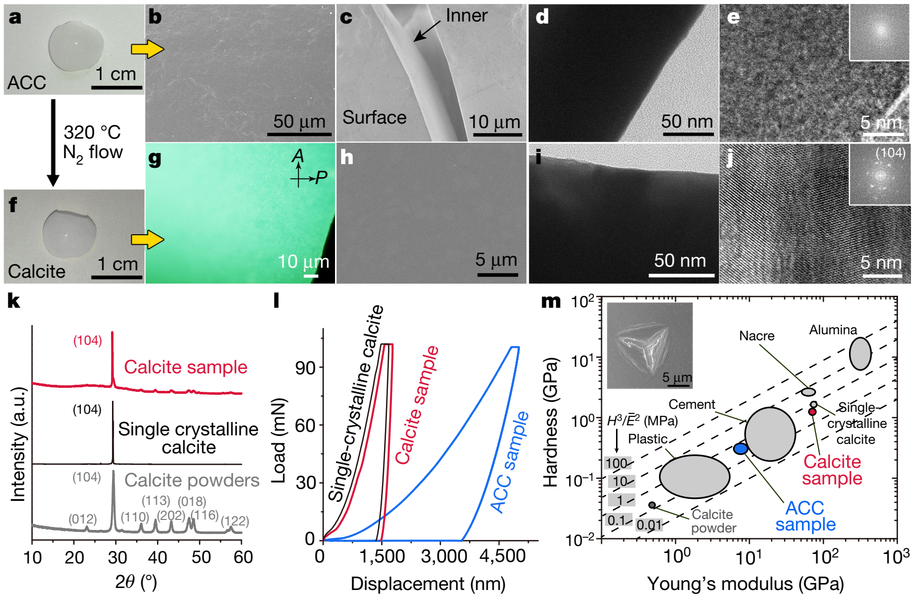


**Fig. 3 | Construction of amorphous and single-crystalline-like $CaCO_3$ bulk materials by the crosslinking of $(CaCO_3)_n$ oligomers.** **a**, Photograph of monolithic ACC prepared from $(CaCO_3)_n$ oligomers. **b**–**e**, SEM (**b**, **c**) and TEM (**d**, **e**) images indicating the continuous solid phase of the prepared monolithic ACC. The inset of **e** is the fast-Fourier-transform image of the sample. Typically, the image of a crack in monolithic ACC exhibits continuity from the surface to the bulk (**c**). **f**, Snapshot of monolithic calcite prepared from monolithic ACC. **g**, Polarized-light optical microscopy (POM) image of the prepared monolithic calcite. **h**, SEM image of a surface on crystallized monolithic $CaCO_3$. **i**, **j**, TEM images of the inner bulk of crystallized monolithic $CaCO_3$. The inset of **j** is the fast-Fourier-transform image of the sample. **k**, XRD pattern of calcite powder, geological single-crystalline calcite and the calcite sample produced from $(CaCO_3)_n$ oligomers. **l**, Load–displacement curves of the ACC sample, calcite sample and geological single-crystalline calcite sample measured by nanoindentation. **m**, Ashby plot of hardness ($H$) against Young's modulus ($E$) for the prepared $CaCO_3$ (including ACC and calcite) and other materials. The upper left inset is an exemplary residual indent of the Berkovich diamond tip on the crystallized $CaCO_3$. $\bar{E}$, plane strain modulus.

interaction between $CO_3^{2-}$ and TEA in gel-like oligomers (Extended Data Fig. 3).

The effect of capping the oligomers with TEA was studied by molecular dynamics simulations (Fig. 2a–c). The stability of the constructed $CaCO_3$ ionic clusters was evaluated by the variation in the Ca–O (from carbonates) coordination number and the cluster size between simulations conducted in the presence and in the absence of TEA. When TEA was not present in the simulation, the Ca–O coordination number quickly increased to 3.2 upon the mixing of $Ca^{2+}$ and $CO_3^{2-}$, and reached a final value of 4.1 after 200 ns. The average cluster size quickly increased to 14 $CaCO_3$ units and reached a final value of 20 units, representing growth of the clusters (Fig. 2a, b). By contrast, in the presence of TEA, the Ca–O coordination number increased slightly and remained at 2.9 after 200 ns. The average cluster size was limited to 2.4 $CaCO_3$ units, demonstrating the formation of $(CaCO_3)_n$ oligomers. The simulation confirms that the linear $(CaCO_3)_n$ oligomers can be stabilized by TEA (Fig. 2c, Extended Data Fig. 4). Notably, a similar chain-like structure for the $CaCO_3$ cluster has been suggested theoretically owing to its relatively low free energy[26]; this is consistent with our findings.

The capped $(CaCO_3)_n$ oligomers are stable in ethanol, as demonstrated by conductivity measurements combined with $^1H$ NMR studies (Extended Data Fig. 5a–c). However, the oligomers are unstable in solvents with high dielectric constants, because the hydrogen bond involved in the capping is weaker in such solvents[23,24] (Extended Data Fig. 5d–i). The volatilization of ethanol removes TEA to initiate crosslinking of the oligomers. In contrast to the reversible aggregation in ethanol solution, the reaction induced by the removal of TEA is irreversible. In situ FTIR spectroscopy revealed that, during the drying of the oligomer gel, both TEA and ethanol were evaporated—as evidenced by a decrease in absorbance at around 1,200 $cm^{-1}$ and at 880 $cm^{-1}$—and calcium carbonate was formed, as demonstrated by an enhancement in the signal at 867 $cm^{-1}$ (Fig. 2d). In situ X-ray diffraction (XRD) studies showed that the resulting mineral phase was amorphous calcium carbonate (ACC) (Extended Data Fig. 6a), and an in situ analysis of the Ca–O coordination number showed that this phase was formed by calcium ions binding to additional carbonate ions through crosslinking (Fig. 2e, Extended Data Fig. 6b). This was evidenced by an increase in the coordination number from around 2.1 (for the oligomer) to around 6.8—the typical Ca–O coordination number of ACC[27] is around 6.7. The crosslinking process during the stepwise growth of $(CaCO_3)_n$ was monitored by transmission electron microscopy (TEM; Fig. 2f, Extended Data Fig. 6c). In this experiment, ultra-dilute $(CaCO_3)_n$ solutions (2.3 mg $l^{-1}$) were sprayed onto TEM grids to avoid aggregation; the increase in the Ca:TEA ratio of the solution mimicked the reduction in TEA capping. The evolution of individual $(CaCO_3)_n$ oligomers could be observed. At Ca:TEA ratios of 1:100, the oligomers were rod-like with lengths of about 1 nm, in agreement with the results from SAXS experiments. At a Ca:TEA ratio of 1:50, the lengths of the rod-like oligomers increased to 3 nm and the growth of a linear chain was observed. Further increase of the Ca:TEA ratio resulted not only in the growth of a linear chain, but also in the formation of a branch structure (Fig. 2f). Our results show an increase in anisotropic chain length (from around 1 nm to around 13 nm) and in the fractal dimension (from around 1.0 to around 1.4) of $(CaCO_3)_n$ upon the removal of TEA (Extended Data Fig. 6d).

The large-scale transformation of the oligomers to larger structures by crosslinking could be tuned by modifying the density of the oligomers (Fig. 2e). When the oligomer density was increased to 11.4 mg $l^{-1}$, the length and width of the resulting product increased and the formation of branches was observed. Energy-dispersive X-ray spectroscopy (EDS) confirmed the composition of the solids as pure $CaCO_3$ (Extended Data Fig. 6e–g). Upon increasing the oligomer density to 22.8 mg $l^{-1}$, the chains became long and thick, and the use of a still higher oligomer density of 57.0 mg $l^{-1}$ resulted in the formation of a network. Further increasing the oligomer density to 285 mg $l^{-1}$ and 1.42 g $l^{-1}$ produced porous and bulk materials, respectively. This evolution from oligomers to chains, networks, and finally bulk materials is similar to phase-separation-based material formation[6].

Centimetre-sized monolithic $CaCO_3$ materials were obtained by the crosslinking of oligomers. The resulting bulk maintained the original

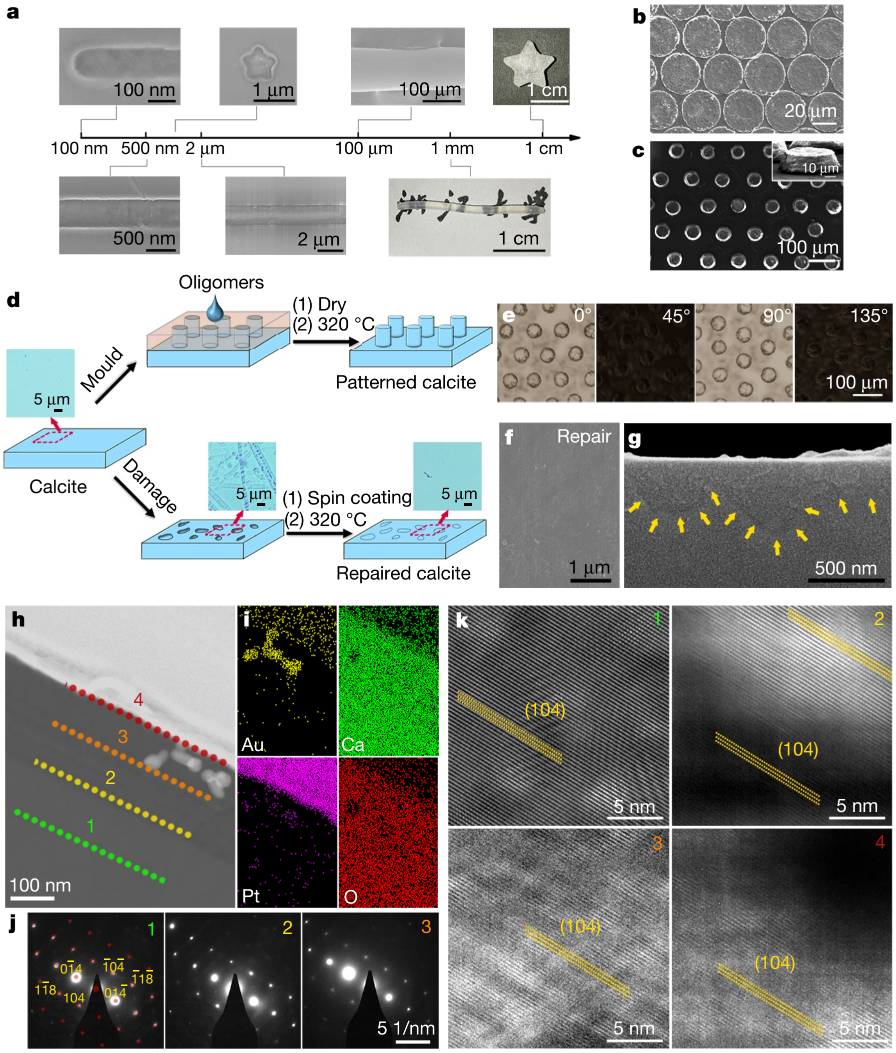


**Fig. 4 | Constructible engineering of CaCO$_3$ single-crystalline materials by using (CaCO$_3$)$_n$ oligomers. a**, Moulded CaCO$_3$ with different dimensions and morphologies. **b**, **c**, Moulded CaCO$_3$ with different patterns. The inset of **c** shows a single CaCO$_3$ rod. **d**, Schemes for pattern construction on single-crystalline calcite (top path), and the repair of rough single-crystalline calcite to smooth calcite (bottom path). The insets show optical microscopy images of the calcite surface at different stages: native, corroded, and repaired. **e**, POM images of the patterned calcite rotated at different angles. **f**, **g**, SEM images of the repaired calcite (surface and cross-section, respectively). **h**, TEM image of a cross-sectional view of the repaired calcite. The different layers labelled 1, 2, 3 and 4 were characterized by selected area electron diffraction and high-resolution lattice fringes in **j** and **k**. **i**, EDS mapping of the repaired calcite in **h**, showing the repaired CaCO$_3$ as well as gold nanolabels. **j**, Selected area electron diffraction patterns of different layers (1–3) of **h** with an aperture of around 170 nm in diameter, showing the same patterns from the bulk to the repaired surface. The red dots in 1 are the simulated diffraction pattern viewed along the $\langle -4, 4, 1 \rangle$ zone axis. **k**, High-resolution lattice fringes at the different layers (1–4) of **h**, exhibiting the facets of (104) with exactly the same orientation from the bulk to the repaired surface.

morphology of the compact gel precursor (Fig. 3a). FTIR spectroscopy and thermal gravimetric analysis indicated the formation of pure ACC without organic residue (Extended Data Fig. 7a, b). Scanning electron microscopy (SEM) and TEM showed the structural continuity in the bulk (Fig. 3b–e), and the internal continuous and integral textures were confirmed by artificially creating a crack (Fig. 3c). At scales from nano- to micrometres, the fabricated material was fully dense and smooth with no porosity or cracks (Fig. 3d, e). A nanoindentation test revealed that the ACC sample (Fig. 3l, blue circle in Fig. 3m) had a Young's modulus of 8.0 ± 1.6 GPa and a hardness of 0.33 ± 0.07 GPa; these values are greater than those of most plastic materials[28]. Furthermore, this preparation of oligomers, as well as their crosslinking, can be extended to other ionic compounds—such as calcium phosphate, cupric phosphate, calcium sulfate and manganous phosphate (Extended Data Fig. 7c–f)—which demonstrates the general applicability of the method.

The crystallization of the ACC bulk could be initiated by either thermal treatment or humidity (Extended Data Fig. 8a, b). The use of an appropriate crystallization gradient resulted in an oriented transformation of the whole bulk with high uniformity (Methods, Extended Data Fig. 9), and the structural continuity in the resulting crystalline phase remained from the nanometre to the centimetre scale (Fig. 3f–j). At the macroscale, the calcite retained the original morphology of the amorphous state (Fig. 3f); at the microscale, there was no particle packing in the crystallized monolith (Fig. 3g, h); and at the nanoscale, the lattice fringes were complete in the bulk (Fig. 3i, j). The strong diffraction on the (104) facet with weak diffraction on the other facets (Fig. 3k) suggested an oriented crystallization at large scale. The profile of a single crystal was further confirmed by a single-crystal diffraction experiment (Extended Data Fig. 8c). Owing to the crystallographic integrity, the calcite monolith exhibited optimal mechanical properties with a Young's modulus and hardness of 71.98 ± 1.32 GPa and 1.42 ± 0.05 GPa, respectively (Fig. 3l, red circle in Fig. 3m)—similar to those of geological single-crystal calcite[29] (Fig. 3l, m).

A considerable advantage of the crosslinking of ionic oligomers is that the oligomeric precursors can be moulded into shapes to enable continuously structured construction (Fig. 4a–c). This in turn enables the engineering of single-crystalline materials, including additive manufacturing. The construction of calcite rod arrays by oligomer crosslinking demonstrates the practicality of the preparation of single-crystal materials with structural complexity (Fig. 4d, e). This method can even be extended to repair damaged single crystals. Calcite single-crystal surfaces in optical devices[30] can be damaged by mechanical crashing, scratching or corrosion, which reduces their functional performance—in particular transmittance. However, $(CaCO_3)_n$ oligomers can generate oriented calcite within nano- and micro-sized pits or ditches of the damaged calcites in order to recover their smooth surface (inset of Fig. 4d, f, g). The repaired region (Fig. 4h, i) had exactly the same crystalline phase and orientation as the bulk beneath (Fig. 4j). The images of the high-resolution lattice fringes from the calcite bulk to the repaired front (Fig. 4k) demonstrate continuous (104) facets without any break, confirming that the same crystalline structure was reproduced exactly, and the transmittance of the repaired single crystal calcite recovered to the same value as that of a pristine sample.

By using humidity- or water-induced crystallization under mild conditions, this method can be extended to the repair of biological hard tissues (biominerals) such as sea-urchin spines and teeth (Extended Data Fig. 10), demonstrating its potential in biological and biomedical applications. The capabilities and advantages of this method result from the properties of the oligomers and their crosslinking, and could enable the production of inorganic materials by a route analogous to that for organic polymers.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1645-x.

Received: 8 February 2018; Accepted: 1 August 2019;
Published online 16 October 2019.

1. Meyers, M. A., Chen, P. Y., Lin, A. Y. M. & Seki, Y. Biological materials: structure and mechanical properties. *Prog. Mater. Sci*. **53**, 1–206 (2008).
2. Riley, F. L. Silicon nitride and related materials. *J. Am. Ceram. Soc*. **83**, 245–265 (2000).
3. Xu, H. H. K. et al. Calcium phosphate cements for bone engineering and their biological properties. *Bone Res*. **5**, 17056 (2017).
4. Hu, C. L. & Li, Z. J. A review on the mechanical properties of cement-based materials measured by nanoindentation. *Constr. Build. Mater*. **90**, 80–90 (2015).
5. Thanh, N. T. K., Maclean, N. & Mahiddine, S. Mechanisms of nucleation and growth of nanoparticles in solution. *Chem. Rev*. **114**, 7610–7630 (2014).
6. Gebauer, D., Kellermeier, M., Gale, J. D., Bergstrom, L. & Colfen, H. Pre-nucleation clusters as solute precursors in crystallisation. *Chem. Soc. Rev*. **43**, 2348–2371 (2014).
7. Gebauer, D., Volkel, A. & Colfen, H. Stable prenucleation calcium carbonate clusters. *Science* **322**, 1819–1822 (2008).
8. Kellermeier, M. et al. Colloidal stabilization of calcium carbonate prenucleation clusters with silica. *Adv. Funct. Mater*. **22**, 4301–4311 (2012).
9. Wolf, S. E. et al. Carbonate-coordinated metal complexes precede the formation of liquid amorphous mineral emulsions of divalent metal carbonates. *Nanoscale* **3**, 1158–1165 (2011).
10. Smeets, P. J. M. et al. A classical view on nonclassical nucleation. *Proc. Natl Acad. Sci. USA* **114**, E7882–E7890 (2017).
11. Gower, L. B. & Odom, D. J. Deposition of calcium carbonate films by a polymer-induced liquid-precursor (PILP) process. *J. Cryst. Growth* **210**, 719–734 (2000).
12. Cheng, X. G. & Gower, L. B. Molding mineral within microporous hydrogels by a polymer-induced liquid-precursor (PILP) process. *Biotechnol. Prog*. **22**, 141–149 (2006).
13. Wolf, S. E. et al. Strong stabilization of amorphous calcium carbonate emulsion by ovalbumin: gaining insight into the mechanism of 'polymer-induced liquid precursor' processes. *J. Am. Chem. Soc*. **133**, 12642–12649 (2011).
14. De Yoreo, J. J. et al. Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* **349**, aaa6760 (2015).
15. Koltzenburg, S., Maskos, M. & Nuyken, O. (eds) *Polymer Chemistry* Ch. 14 (Springer, 2017).
16. Lewis, J. A. Colloidal processing of ceramics. *J. Am. Ceram. Soc*. **83**, 2341–2359 (2000).
17. Eckel, Z. C. et al. Additive manufacturing of polymer-derived ceramics. *Science* **351**, 58–62 (2016).
18. Liu, H. L., Gong, Q. H., Yue, Y. H., Guo, L. & Wang, X. Sub-1 nm nanowire based superlattice showing high strength and low modulus. *J. Am. Chem. Soc*. **139**, 8579–8585 (2017).
19. Truby, R. L. & Lewis, J. A. Printing soft matter in three dimensions. *Nature* **540**, 371–378 (2016).
20. Feinle, A., Elsaesser, M. S. & Husing, N. Sol-gel synthesis of monolithic materials with hierarchical porosity. *Chem. Soc. Rev*. **45**, 3377–3399 (2016).
21. Caminade, A. M., Hey-Hawkins, E. & Manners, I. Smart inorganic polymers. *Chem. Soc. Rev*. **45**, 5144–5146 (2016).
22. Drake, K., Mukherjee, I., Mirza, K., Ji, H. F. & Wei, Y. Phenylethynyl and phenol end-capping studies of polybiphenyloxydiphenylsilanes for cross-linking and enhanced thermal stability. *Macromolecules* **44**, 4107–4115 (2011).
23. Sautter, A., Thalacker, C., Heise, B. & Wurthner, F. Hydrogen bond-directed aggregation of diazadibenzoperylene dyes in low-polarity solvents and the solid state. *Proc. Natl Acad. Sci. USA* **99**, 4993–4996 (2002).
24. Liu, Z. M. et al. Ionization controls for biomineralization-inspired $CO_2$ chemical looping at constant room temperature. *Phys. Chem. Chem. Phys*. **17**, 10080–10085 (2015).
25. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr*. **42**, 342–346 (2009).
26. Demichelis, R., Raiteri, P., Gale, J. D., Quigley, D. & Gebauer, D. Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nat. Commun*. **2**, 590 (2011).
27. Michel, F. M. et al. Structural characteristics of synthetic amorphous calcium carbonate. *Chem. Mater*. **20**, 4720–4728 (2008).
28. Weaver, J. C. et al. Analysis of an ultra hard magnetic biomineral in chiton radular teeth. *Mater. Today* **13**, 42–52 (2010).
29. Broz, M. E., Cook, R. F. & Whitney, D. L. Microhardness, toughness, and modulus of Mohs scale minerals. *Am. Mineral*. **91**, 135–142 (2006).
30. Hirano, S., Yogo, T. & Kikuta, K. Synthetic calcite single-crystals for optical-device. *Prog. Cryst. Growth Charact. Mater*. **23**, 341–367 (1992).

## METHODS

**Reagents.** Calcium chloride dihydrate ($CaCl_2{\cdot}2H_2O$, powder, 99.0%, Sigma-Aldrich), triethylamine ($(C_2H_5)_3N$, 99%, Sigma-Aldrich), ethanol ($C_2H_5OH$, 99.7%, Aladdin), cupric chloride dihydrate ($CuCl_2{\cdot}2H_2O$, powder, ACS reagent 99%, Sigma-Aldrich), manganous chloride tetrahydrate ($MnCl_2{\cdot}4H_2O$, powder, ACS reagent 98%, Sigma-Aldrich), sulfuric acid ($H_2SO_4$, 98%, AR, Sinopharm), phosphoric acid ($H_3PO_4$, 99% powder, Aladdin), ethanol-$d_6$ ($C_2D_5OD$, 99.5%, Aldrich), DMSO-$d_6$ ($(CD_3)_2SO$, 100%, Aldrich) and water-$d_2$ ($D_2O$, 100%, Aladdin) were used as received.

**Preparation of ionic oligomers.** First, 0.05–0.40 g $CaCl_2{\cdot}2H_2O$ was dissolved in 100 ml ethanol, and 0.238–7.6 ml TEA was added to prepare a mixed solution. Second, 100 ml $\min^{-1}$ of $CO_2$ was bubbled into the solution while stirring, and the solution became turbid after few minutes. Finally, the $CO_2$ was kept bubbling for 10 min, and the solution was stirred for another 30 min to obtain a transparent or semi-transparent solution that contained $(CaCO_3)_n$ oligomers, which was named as the original $(CaCO_3)_n$ oligomer solution.

The $(CaCO_3)_n$ oligomers were concentrated by centrifugation (8,000–27,500*g*, Allegra 64R Centrifuge, Beckman Coulter) and redispersed in ethanol twice to remove impurities. The use of a higher centrifugal force can increase the concentration but makes it harder to redisperse the oligomers. The $(CaCO_3)_n$ oligomers were characterized or used in two states: oligomers were redispersed into ethanol ($(CaCO_3)_n$ oligomer solution); or oligomers were centrifuged in a gel state (gel-like $(CaCO_3)_n$ oligomers).

The preparation of calcium phosphate, cupric phosphate, calcium sulfate and manganous phosphate was similar. Typically, 60 ml ethanol solution containing 25 mM cation ($Ca^{2+}$, $Cu^{2+}$ or $Mn^{2+}$) and 40 ml ethanol solution containing 25 mM $H_3PO_4$ or 37.5 mM $H_2SO_4$ with 750 mM TEA were prepared. The subsequent quick mixing of these two solutions by stirring generates ionic oligomers within 30 min. All gel-like oligomers can be collected by the high-speed centrifugation (27,500*g*) of their suspensions.

**Mass spectrometry.** Original $(CaCO_3)_n$ oligomer solution, which was prepared from 0.05g $CaCl_2{\cdot}2H_2O$ with different ratios of TEA, was directly used for analysis by mass spectrometry. A Bruker amaZon ion trap mass spectrometer (Bruker-Franzen Analytik) that was equipped with an electrospray ionization (ESI) source interface in positive and negative ion modes was used. Nitrogen was used as the nebulization gas at a pressure of 10 psi and the drying gas at a flow rate of 5 l $\min^{-1}$. The drying gas temperature was set to 250 °C, and the capillary voltage was set to ±4,500 V with a scan speed of 8,100 $m/z$ per second in enhanced resolution mode. The $(CaCO_3)_n$ oligomer suspension was infused into the mass spectrometer with a syringe pump at a flow rate of 180 μl $h^{-1}$. The mass spectra were obtained with helium as the collision gas at an appropriate collision energy. All the presented data was corrected according to background data.

**Nuclear magnetic resonance spectroscopy.** First, 1 ml of freshly prepared $(CaCO_3)_n$ oligomer solutions were centrifuged and redispersed into 1 ml ethanol-$d_6$, DMSO-$d_6$, and water-$d_2$. Second, these solutions were sealed in NMR tubes and immediately measured on a 600 MHz Direct Drive 2 NMR spectrometer (Agilent) at room temperature for the $^1H$ spectra. Owing to the uploading and measurement time, these groups were considered as the '30 min groups' in Extended Data Fig. 5. Third, the sample was aged for one day (two days for the ethanol group) and measured again to obtain the 'one day group' data. With the exception of the $^1H$ (on the ethyl group) shift for TEA, the shift of $^1H$ (on both ethyl group and hydroxyl group) on ethanol could not be observed, which indicates that ethanol had a minimal influence on $(CaCO_3)_n$ oligomers (Extended Data Fig. 5j–l). The $^{13}C$ NMR experiments were performed using the original $(CaCO_3)_n$ oligomer solution at a concentration of 11.4 mg $l^{-1}$ with different $CaCl_2$:TEA molar ratios of 1:5, 1:20, 1:50 and 1:100.

**Synchrotron small-angle X-ray scattering.** SAXS data were obtained on the BL19U2 beamline at the Shanghai Synchrotron Radiation Facility. Scattered X-ray intensities were measured by a Pilatus 1 M detector (Dectris) at the X-radiation wavelength ($\lambda$) of 0.918 Å. The distance of the sample to the detector was set at 2,111.0 mm for the measurements. Each sample was measured 20 times with a 1-s exposure time for each. The solutions were prepared as for the mass spectrometry studies. Different concentrations of $(CaCO_3)_n$ oligomer solution (11.4 mg $l^{-1}$, 57.0 mg $l^{-1}$ and 285 mg $l^{-1}$) with Ca:TEA = 1:100 was measured. The blank groups were ethanol solution with TEA (the concentration of TEA is determined by the related sample). In the reversible aggregation–disaggregation experiment using 285 mg $l^{-1}$ oligomer ethanol solution, the dilution was performed by adding ethanol and sonicating for 2 h at room temperature. The condensation was performed by evaporation in vacuum conditions. The acquired raw data were analysed by the software package BioXTAS RAW 1.2.1[31]. Then, the background subtracted data were analysed by ATSAS[32] for the representative reconstruction of the three-dimensional shape of the samples. The ab initio simulation was performed using the DAMMIN software package[25]. The shapes of the samples at different concentrations were determined using DAMMIN (average with DAMAVER and refine with DAMMIN). The length of sample was read from the simulated 'Approx. shape dimension ($r_o$)'.

**Dynamic light scattering.** The hydrodynamic diameter distributions of $(CaCO_3)_n$ oligomers and aggregates were examined by a dynamic light scattering photometer (ZEN 3600, Malvern Instruments). Different concentrations of $(CaCO_3)_n$ oligomers at a ratio of $CaCl_2$:TEA = 1:20 were measured at 25 °C. Typically, more than 2 h sonication at room temperature is necessary in order to disaggregate the oligomer aggregates.

**Radial distribution function and coordination number.** The $(CaCO_3)_n$ oligomer suspension (sample) and ethanol (background) were prepared on a quartz plate for X-ray diffraction measurements with an Ultimate-IV diffractometer (Rigaku). The incident X-ray wavelength was 0.154 nm, and the two-theta degree for measurements were operated from 5–140° with a scanning speed of 15° $\min^{-1}$. The change in coordination number during crosslinking was measured in situ using gel-like oligomers as samples. Each measurement was operated at a scanning speed of 20° $\min^{-1}$ with subsequent 20 min-intervals for the next measurement. The radial distribution function (RDF) was calculated from the diffraction data using PDFgetX3[33]. All parameters were established according to the instructions. The peak at around 2.4 Å in the RDF was contributed by the first coordination shell of Ca–O[27], and the coordination number of Ca–O was calculated by the following equation:

$$n(r) = 4\pi\rho \int_{r_{\min}}^{r_{\max}} g(r) r^2 \mathrm{d}r$$

where the $g(r)$ (RDF) peak located between $r_{\min}$ and $r_{\max}$ corresponds to the first Ca–O coordination shell; and $\rho$ is the number density, which was 0.094 and 0.049 for ethanol and $(CaCO_3)_n$, respectively. Owing to the relatively low concentration of $(CaCO_3)_n$ oligomers in suspension, a $\rho$ of 0.094 was chosen and the coordination number of Ca–O in the oligomers was roughly 2. Owing to the uncertainty of an exact $\rho$ during the crosslinking process, the 'maximum' and 'minimum' coordination numbers were also presented in Fig. 2e, by assuming the content of ethanol was 100% and 0%, respectively.

**Molecular dynamics simulation.** All the molecular dynamics simulations in this work are performed in the NpT ensemble using the Gromacs 4.6.7 package[34,35]. The visualization was generated via visual molecular dynamics[36]. Periodic boundary condition was applied to all the three directions. A time step of 1 fs was used with atom coordinates saved every 10 ps. A temperature of 298 K was maintained by the Nosé–Hoover thermostat with a 0.1 ps relaxation time, and a pressure of 1 atm was maintained using the Berendsen method[37]. The particle mesh Ewald summation[38] was used to calculate the long-range electrostatic interaction, with a cutoff of 1.3 nm for the separation of the direct and reciprocal space summation. The cutoff distance for the van der Waals interaction was 1.3 nm. For the force-field parameters of calcite, a previously developed model was used (Lennard-Jones (12, 6) potentials)[39], which is fitted from another model (Buckingham potentials)[40]. The optimized potentials for liquid simulations (OPLS) all-atom force field was used to describe the ethanol molecules and the trimethylamine molecules[41].

In the molecular dynamics simulations, two systems with different numbers of TEA molecules (0 and 400) were constituted by around 81,700 atoms in a box of $9.54 \times 9.54 \times 9.54\ \mathrm{nm}^3$, including 400 $Ca^{2+}$, 400 $CO_3^{2-}$, and around 7,000 ethanol molecules. Three independent parallel simulations (200-ns run) were carried out for each of the systems to obtain reliable results. The cluster analysis was carried out by the *g_clustsize* tool in the Gromacs package.

**Fourier transform infrared spectroscopy.** Gel-like $(CaCO_3)_n$ oligomers were prepared between two sealed KBr plates for the measurements. The FTIR spectra were measured using an IRAffinity-1 infrared spectrometer (Shimadzu) with 32 scans at a resolution of 2 $cm^{-1}$. The backgrounds were determined using blank KBr plates. The gel-like $(CaCO_3)_n$ oligomers, which mainly consisted of ethyl acetate, were produced by washing gel-like $(CaCO_3)_n$ oligomers in ethyl acetate twice; after that, the ethanol in gel was replaced by ethyl acetate.

The crosslinking of oligomers were measured in situ by using the GS10800-B Quest sampling accessory (Specac) with a diamond attenuated total reflectance (ATR) sampling plate. Gel-like $(CaCO_3)_n$ oligomers were placed on the sampling plate and sealed in a volatiles cover for FTIR measurements with 32 scans at a resolution of 2 $cm^{-1}$. The volatiles cover was removed between each measurement for 2 min to promote the volatilization of TEA and ethanol. The blank sampling plate was considered as the background.

**Liquid-cell transmission electron microscope.** Liquid cell chips (Hummingbird Scientific), which were similar to those previously described[42,43], were used for the liquid-cell TEM observation. The liquid cell consists of two square 4 $mm^2$ silicon chips with 50-nm-thick silicon nitride membranes in $50 \times 200\ \mu m^2$ windows for imaging. All chips were plasma-cleaned in a Plasma Cleaner (Chendu Mingheng Science & Technology) for 5 min before use by bleeding in 200–400 mTorr of ambient atmosphere. $(CaCO_3)_n$ oligomers (57.0 mg $l^{-1}$ and 285 mg $l^{-1}$) were redispersed into isopropanol and dropped onto chips for observation (Extended Data Fig. 2). Electron microscopy was conducted in a field-emission FEI Tecnai G$^2$ F20 (FEI) operated at 200 kV. The TEM images were captured by an Orius Gatan camera. Snapshots were recorded using the software Bandicam (Bandisoft) and were further processed using ImageJ.

**Viscosity measurements.** Approximately 25 ml of gel-like $(CaCO_3)_n$ with around 20 wt% inorganic composition was obtained. Then, the gel was dispersed into different volumes of ethanol to prepare solutions that contain 39.5, 27.6, 15.8, 7.9 and 3.9 g $l^{-1}$ of oligomers. The solutions were cooled at 5 °C for one hour before measurement. The viscosity was measured using an NDJ-8S rotational viscometer (Bangxi Instrumental) with a #1 rotator at the speed range of 3–6 r.p.m. The Huggins equation and the Einstein equation were used to fit the measured data[44]. The viscosity of chain-like structures can be fitted using the Huggins equation:

$$\frac{\eta_{sp}}{C} = [\eta] + K[\eta]^2 C$$

where $\eta_{sp}$, $[\eta]$, $K$ and $C$ are the specific viscosity, the intrinsic viscosity, a proportionality constant, and the concentration of the oligomers, respectively. By contrast, the viscosity of spherical particles in solution is fitted using the Einstein equation:

$$\frac{\eta_{sp}}{C} = \frac{K}{\rho}$$

where $K$ is a constant and $\rho$ is the particle density.

According to the fittings, the $R^2$ for the Huggins equation and the Einstein equation were 0.98 and 0.94, respectively. Thus, the aggregates of $(CaCO_3)_n$ oligomers are most similar to a chain-like structure.

**Conductivity measurements.** First, 50 ml of freshly prepared $(CaCO_3)_n$ oligomer suspensions was centrifuged, and the obtained gel was redispersed in 2.5 ml ethanol. Second, 0.5 ml of the solution was injected into 9.5 ml ethanol, DMSO or water. The solution conductivities were subsequently measured with a conductivity meter (DDSJ-308A, INESA Scientific Instrument) at 18 °C.

**TEM observation of crosslinking oligomers.** A technical improvement is applied to ensure the direct observation of individual oligomers and their growth under high-resolution TEM. The improvement includes spraying ultra-diluted oligomer solution to avoid species aggregation, integrating multiple images, and enhancing image contrast by bandpass to improve resolution. In order to observe an individual oligomer without any aggregation, ultra-dilute $(CaCO_3)_n$ solutions (2.3 mg $l^{-1}$) with different Ca:TEA ratios were prepared and they were sprayed on plasma-cleaned ultra-thin carbon support films, ensuring the isolation of the oligomers. The electron microscopy was conducted on a field-emission FEI Tecnai G2 F20 (FEI) operated at 200 kV. The decrease in the amount of TEA (increase of Ca:TEA ratio) mimicked the reduction of the TEA capping. Multiple images were acquired and they were integrated by ImageJ to increase the signal-to-noise ratio. The contrasts of samples were enhanced by using bandpass filter in ImageJ[45]. The lengths and the fractal dimensions of the observed species under TEM were also measured and analysed by ImageJ.

In order to observe a large-scale evolution from $(CaCO_3)_n$ to bulk materials under TEM, different concentrations of as prepared $(CaCO_3)_n$ oligomer solution (11.4 mg $l^{-1}$ to 1.42 g $l^{-1}$) were dried on an ultra-thin carbon support film and the electron microscopy was conducted on a HT-7700 (Hitachi) TEM operated at 100 kV.

**Preparation of monolithic materials.** The gel-like ionic oligomers (such as $(CaCO_3)_n$) were prepared by high-speed centrifugation (27,500*g*). Direct drying of the gel can cause the formation of monolithic materials. Vacuum can promote the drying rate but increases the possibility of generating cracks. Typically, obtaining the monolith by drying at 0.04 MPa while minimizing cracks was an efficient process. Owing to the liquid-like property of gel-like oligomers, the gels were moulded in different moulds before drying. The microstructure silicon moulds were prepared by using micro-processing of Quanta 3D FEG focused ion beam (FIB, Carl Zeiss). All moulded gels had to be compacted or the oligomers were unable to crosslink, which caused the formation of fractures. Electron microscopy was conducted in a field-emission FEI Tecnai G$^2$ F20 (FEI) operated at 200 kV and a field-emission SU-8010 (Hitachi) operated at 5 kV.

**Thermally induced ACC crystallizations.** The amorphous phase is commonly considered as a thermodynamically unstable phase, but its transformation into the crystalline phase is slow at room temperature in the absence of water[46,47]. Generally, thermal treatment is applied to accelerate this crystallization process[48,49]. Typically, the monolithic amorphous $CaCO_3$ could be transformed to calcite by the following treatment. First, the sample was put onto a heating stage and protected by $N_2$ gas (to keep dry). Second, the stage was heated to 320 °C at a rate of 1 °C $min^{-1}$. Third, the sample was maintained at 320 °C for 30 min. Finally, the sample was cooled at a rate of 5 °C $min^{-1}$. After such thermal treatment, the volume of the resulting calcite crystal was around 94% of the original ACC bulk. The success rate in generating non-cracked single calcite crystals was dependent upon the bulk sizes. At scale of a hundred micrometres, the rate could reach almost 100%; at scale of a few centimetres, the rate was about 30%; however, we think that this value can be increased by future technical improvements.

By using different thermal treatment protocols, we found that the formation of single-crystal-like $CaCO_3$ was sensitive to heating rate ($H$), temperature ($T$) and cooling rate ($C$) because the thermal gradient in the calcium carbonate bulk had an important role in controlling the oriented crystallization frontier (Extended Data Fig. 9). The infrared thermal images were acquired by FLIR ONE Pro thermal camera (FLIR Systems) during the processing of thermal treatment protocol. The thermal images indicated that a well-organized thermal gradient diffused from the proximal heating site to the distal site (Extended Data Fig. 9a). Meanwhile, the initial solid-state transition from ACC to calcite occurred at the proximal heating site and then diffused towards the distal site and the key controls were: (1) $T$. A lower temperature (300 °C) could not induce crystallization (Extended Data Fig. 9e), but a higher temperature (325 °C) could destroy the thermal gradient for oriented crystallization and result in polycrystal formation (Extended Data Fig. 9f); (2) $H$. A slow heating rate benefitted the establishment of a well-organized thermal gradient in the crystallization bulk. At an increased $H$ of 2 °C $min^{-1}$, random crystallization occurred owing to the break in thermal gradient (Extended Data Fig. 9g); (3) $C$. A slow cooling rate benefitted the uniform shrinkage. At an increased $C$ of 10 °C $min^{-1}$, cracks could form readily in the crystal bulk (Extended Data Fig. 9h); (4) An insufficient temperature maintenance time at 320 °C led to incomplete crystallization (Extended Data Fig. 9c); (5) The particle-packing ACC bulk only resulted in polycrystals, owing to its discontinuous structure (Extended Data Fig. 9i).

The resulting single-crystal sample was used for the X-ray single-crystal diffraction test (Oxford Diffraction Gemini-A-Ultra with Atlas CCD and graphite-monochromated Mo K$\alpha$ radiation). The indexing rate of the sample was 79% (Extended Data Fig. 8c), indicating that the sample is reliable for the collection of single-crystal diffraction data. For a comparison, the indexing rate of the geological single-crystalline calcite was 88%.

**Humidity-induced ACC crystallization.** ACC is unstable under conditions of high relative humidity because water can readily induce its crystallization even at room temperature[46]. The humidity-induced crystallization was performed by incubating the monolithic ACC bulks with a thermostatic water bath at a temperature of 25 °C, in which the relative humidity was measured to be 100%. The samples were periodically examined by FTIR (IRAffinity-1 infrared spectrometer, Shimadzu) (Extended Data Fig. 8a, b). The ratio of peak areas at 875 $cm^{-1}$ and 866 $cm^{-1}$ was used to quantify the bulk crystallinity[50].

**Nanoidentation measurements.** A Nano Indentation (Agilent G200) with a Berkovich diamond probe was used to perform the nanoindentation tests on the prepared $CaCO_3$ samples and single-crystalline calcite (geological Iceland spars). To ensure that the indentations were performed on plaint areas, more than 10 indentation locations were manually selected under an optical microscope at a magnification factor of 1,000. The maximum load was 100 mN and Poisson's ratio was 0.28 according to previous studies[51].

**Construction of calcite patterns.** First, the moulds were fixed on the surface of single-crystalline calcite or glass by using cyanoacrylate-type adhesives. Typically, we chose TEM copper grids with different sized holes as the moulds. The gel-like $(CaCO_3)_n$ oligomers were prepared by high-speed centrifugation (27,500*g*), and then the freshly prepared gel was squeezed into the holes (any overflowing gel was removed). The gel was dried at room temperature within a few minutes to obtain the ACC phase. After that, the sample was immersed in acetone to remove the cyanoacrylate-type adhesives. Then, the mould could be removed, and the $CaCO_3$ rod or film arrays were obtained. To construct a calcite array, the as-prepared sample was then heated to 320 °C at a ramp rate of 1 °C $min^{-1}$ under an $N_2$ atmosphere, followed by another 30 min of incubation at 320 °C. Finally, the calcite array was built by cooling the sample to room temperature.

**Repair of damaged calcite single crystal.** Geological Iceland spars were used as the single-crystalline calcite model, and they were washed with water twice before use. The damage was produced by dropping 16 mM HCl/ethanol solution on the surface of calcite followed by 30 s of incubation. After that, the acidic solution was directly wiped off the surface, and nano- and micro-sized pits and ditches were generated. For the repair of damaged calcite, 40 ml of freshly prepared $(CaCO_3)_n$ oligomer solution ($CaCl_2$:TEA = 1:20) was centrifuged at 10,000*g*, and the $(CaCO_3)_n$ oligomers were redispersed into 3 ml of ethanol. At the same time, 0.5 ml of 10 μM Au nanorod[52] solution was added to the solution for labelling. The solution was sonicated at room temperature for 1 h. Then, around 100 μl of sonicated $(CaCO_3)_n$ oligomer solution was spin-coated onto the damaged surface of calcite at a rotation speed of 800 r.p.m. Subsequently, an amorphous calcium carbonate repair layer covered the calcite surface. To recover the crystallinity, the calcite was heated to 320 °C at a ramp rate of 1 °C $min^{-1}$ under $N_2$ atmosphere, followed by another 10 min of incubation at 320 °C. Finally, the repair was finished by cooling the calcite to room temperature.

The cross-sectional slice of repaired calcite was prepared by micro-processing with a Quanta 3D FEG focused ion beam (FIB, Carl Zeiss). Electron microscopy was conducted on a field-emission FEI Tecnai $G^2$ F20 microscope (FEI) operated at 200 kV and a field-emission SU-8010 microscope (Hitachi) operated at 5 kV. The relative transmittance was measured by using an Eclipse 80i (Nikon) and ImageJ. The same calcite in different states (native, damaged and repair) was observed in transparent mode. The images were acquired under exactly the same observation conditions by using NIS-Elements F 3.0, and the intensities ($I_n$ for native, $I_d$ for damaged and $I_r$ for repair) of the images were measured by ImageJ. As the transmittance $= I/I_0$, where $I_0$ is the incident light intensity, the relative transmittance can be calculated by using the transmittance of native calcite as a control. Thus, the relative transmittance $= I_d/I_n$ for damaged calcite, and the relative transmittance $= I_r/I_n$ for repaired calcite. By this method, the relative transmittance of the damaged calcite could be recovered from around 0.6 to around 1.0.

**Sea urchin spine repair and enamel repair.** The sea-urchin spines were cut to a length of around 0.5 cm, and the removal of external organic matter was performed by dispersing the spines pieces in 3% NaOCl solution for 1 h[53]. The spines were then rinsed with water and air dried. After that, the $(CaCO_3)_n$ oligomer gel was coated onto the spine. After the ACC formation on the spine by air-drying at room temperature (25 °C), the sample was immersed into simulated sea water (containing 26 g $l^{-1}$ NaCl and 24 g $l^{-1}$ $MgCl_2$) to induce crystallization at room temperature. The sample was cleaned by sonication before examination by SEM. The natural and repaired sea-urchin spines were examined by FTIR spectroscopy (IRAffinity-1 infrared spectrometer, Shimadzu) on a diamond ATR sampling plate. For conventional solution crystallization, the cleaned and dried spine was immersed into the above mentioned simulated sea water containing 20 mM $CaCl_2$. The crystallization occurred by a slow diffusion of $NH_3$ and $CO_2$ (from $(NH_4)_2CO_3$ powder) into the solution.

Enamel windows (provided by Zhejiang University Hospital with an approval from the Ethics Committee of Zhejiang University) were etched with 37 wt% $H_3PO_4$ for 30 s, ultrasonicated in deionized water for 20 min and air-dried. About 50 μl of 4.2 g $l^{-1}$ calcium phosphate oligomers were dropped onto the enamel surface and then air-dried. The crystallization was initiated by immersing the sample into a simulated oral fluid[54] containing 1.5 mM $CaCl_2$ and 0.9 mM $K_2HPO_4$ at 37 °C for 1 day.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

31. Nielsen, S. S. et al. BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J. Appl. Crystallogr.* **42**, 959–964 (2009).
32. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).
33. Juhás, P., Davis, T., Farrow, C. L. & Billinge, S. J. L. PDFgetX3: a rapid and highly automatable program for processing powder diffraction data into total scattering pair distribution functions. *J. Appl. Crystallogr.* **46**, 560–566 (2013).
34. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
35. Van der Spoel, D. et al. *Gromacs User Manual version 4.5* (2010).
36. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
37. Berendsen, H. J., Postma, J. V., van Gunsteren, W. F., DiNola, A. & Haak, J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
38. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N*·log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
39. Shen, J. W., Li, C., van der Vegt, N. F. & Peter, C. Understanding the control of mineralization by polyelectrolyte additives: simulation of preferential binding to calcite surfaces. *J. Phys. Chem. C* **117**, 6904–6913 (2013).
40. Raiteri, P., Gale, J. D., Quigley, D. & Rodger, P. M. Derivation of an accurate force-field for simulating the growth of calcium carbonate from aqueous solution: a new model for the calcite–water interface. *J. Phys. Chem. C* **114**, 5997–6010 (2010).
41. Jorgensen, W. L. & Tiradorives, J. The Opls potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).
42. Nielsen, M. H., Aloni, S. & De Yoreo, J. J. In situ TEM imaging of $CaCO_3$ nucleation reveals coexistence of direct and indirect pathways. *Science* **345**, 1158–1162 (2014).
43. Smeets, P. J., Cho, K. R., Kempen, R. G., Sommerdijk, N. A. & De Yoreo, J. J. Calcium carbonate nucleation driven by ion binding in a biomimetic matrix revealed by in situ electron microscopy. *Nat. Mater.* **14**, 394–399 (2015).
44. Brinker, C. J. & Scherer, G. W. Sol-gel-glass. 1. Gelation and gel structure. *J. Non-Cryst. Solids* **70**, 301–322 (1985).
45. Oshida, K. et al. Analysis of pore structure of activated carbon-fibers using high-resolution transmission electron-microscopy and image-processing. *J. Mater. Res.* **10**, 2507–2517 (1995).
46. Ihli, J., Kulak, A. N. & Meldrum, F. C. Freeze-drying yields stable and pure amorphous calcium carbonate (ACC). *Chem. Commun.* **49**, 3134–3136 (2013).
47. Konrad, F., Gallien, F., Gerard, D. E. & Dietzel, M. Transformation of amorphous calcium carbonate in air. *Cryst. Growth Des.* **16**, 6310–6317 (2016).
48. Ihli, J. et al. Dehydration and crystallization of amorphous calcium carbonate in solution and in air. *Nat. Commun.* **5**, 3169 (2014).
49. Xu, X., Han, J. T., Kim, D. H. & Cho, K. Two modes of transformation of amorphous calcium carbonate films in air. *J. Phys. Chem. B* **110**, 2764–2770 (2006).
50. Tao, J. H., Zhou, D. M., Zhang, Z. S., Xu, X. R. & Tang, R. K. Magnesium-aspartate-based crystallization switch inspired from shell molt of crustacean. *Proc. Natl Acad. Sci. USA* **106**, 22096–22101 (2009).
51. Kim, Y. Y. et al. An artificial biomineral formed by incorporation of copolymer micelles in calcite crystals. *Nat. Mater.* **10**, 890–896 (2011).
52. Sánchez-Iglesias, A. et al. High-yield seeded growth of monodisperse pentatwinned gold nanoparticles through thermally induced seed twinning. *J. Am. Chem. Soc.* **139**, 107–110 (2017).
53. Politi, Y., Arad, T., Klein, E., Weiner, S. & Addadi, L. Sea urchin spine calcite forms via a transient amorphous calcium carbonate phase. *Science* **306**, 1161–1164 (2004).
54. Li, L. et al. Bio-inspired enamel repair via Glu-directed assembly of apatite nanoparticles: an approach to biomaterials with optimal characteristics. *Adv. Mater.* **23**, 4695–4701 (2011).
55. Koeppe, B. et al. Polar solvent fluctuations drive proton transfer in hydrogen bonded complexes of carboxylic acid with pyridines: NMR, IR and ab initio MD study. *Phys. Chem. Chem. Phys.* **19**, 1010–1028 (2017).
56. Shawoosh, A. S. & Kutub, A. A. An investigation of the electrical, optical and DSC properties of a copper phosphate-glass composition. *J. Mater. Sci.* **28**, 5060–5064 (1993).

**Acknowledgements** We thank J. J. De Yoreo, H. Cölfen, P. Fratzl, N. A. J. M. Sommerdijk, D. Joester and L. B. Gower for discussions; Y. Li and Y. Qiu for help with SAXS data analysis; C. Yang and B. Wang for assistance with synchrotron SAXS at the Shanghai Synchrotron Radiation Facility; C. Jin, F. Chen, W. Wang and Y. Wang for assistance with electron microscopy; J. Liu for assistance with XRD; and M. Yu and Y. Liu for assistance with NMR. This work was supported by the National Natural Science Foundation of China (21625105 and 21805241) and the China Postdoctoral Science Foundation (2017M621909 and 2018T110585). We thank W. Liu for his help and inspiration.

**Author contributions** Z.L. and R.T. initiated the project. Z.L. performed the syntheses and the FTIR, MS, NMR and XRD experiments, the calcite repair and conductivity experiments; C.S. carried out the nanoindentation and SEM experiments; B.J. performed the TEM experiments; Z.Z. performed the computer simulations; Y.Z. acquired the synchrotron SAXS and SEM data; R.T. and Z.L. supervised and supported the project; and Z.L. and X.X. analysed the data. The manuscript was written by Z.L. and R.T. All authors reviewed and approved the manuscript.

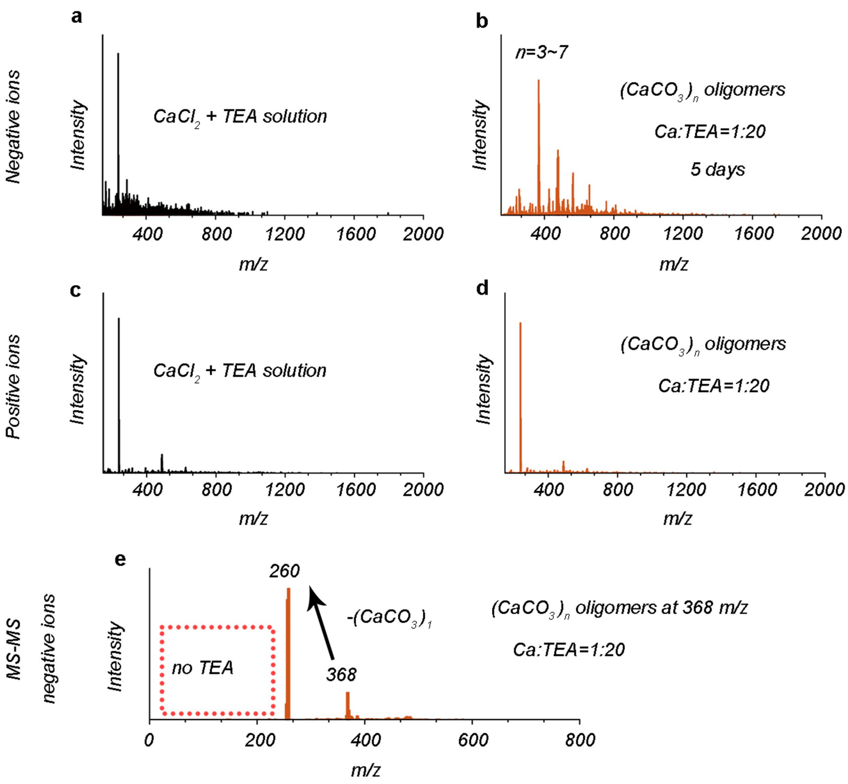
**Competing interests** The authors declare no competing interests.

**Additional information**

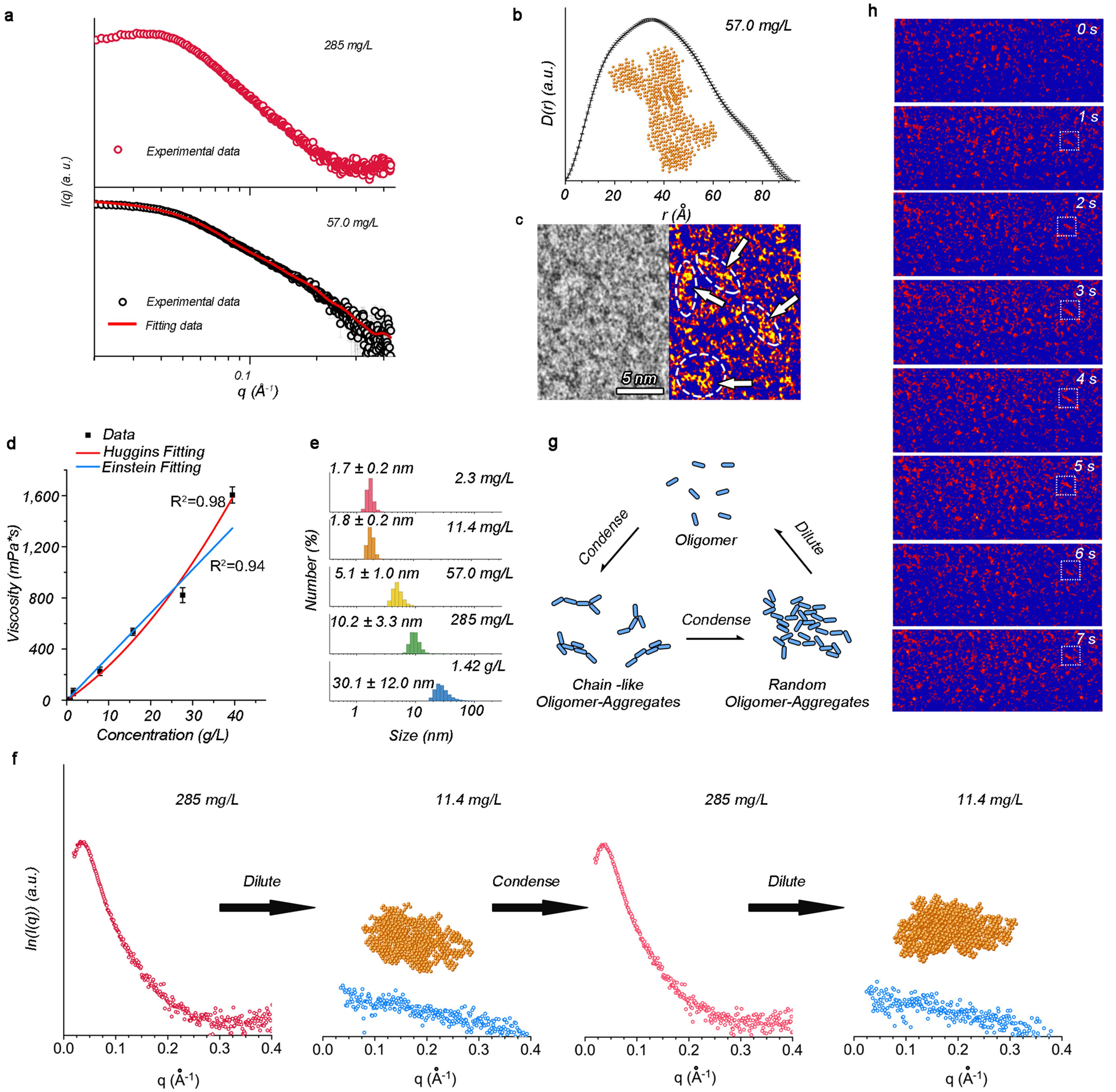
**Correspondence and requests for materials** should be addressed to R.T.

**Peer review information** *Nature* thanks Kazuo Onuma and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

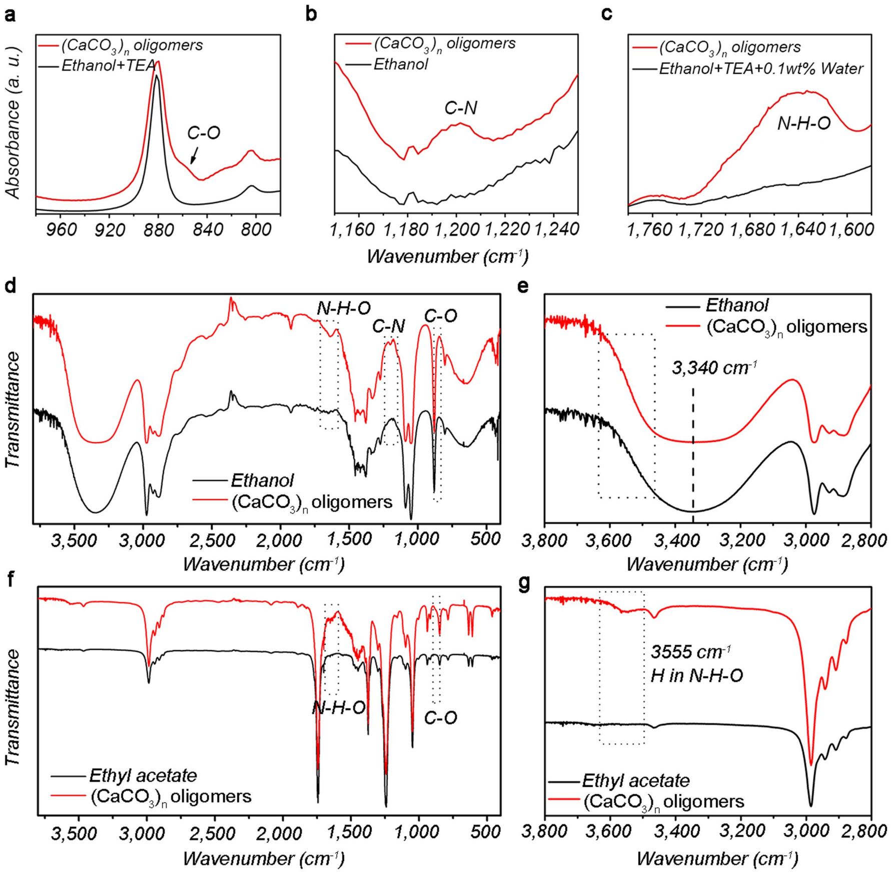
**Reprints and permissions information** is available at http://www.nature.com/reprints.



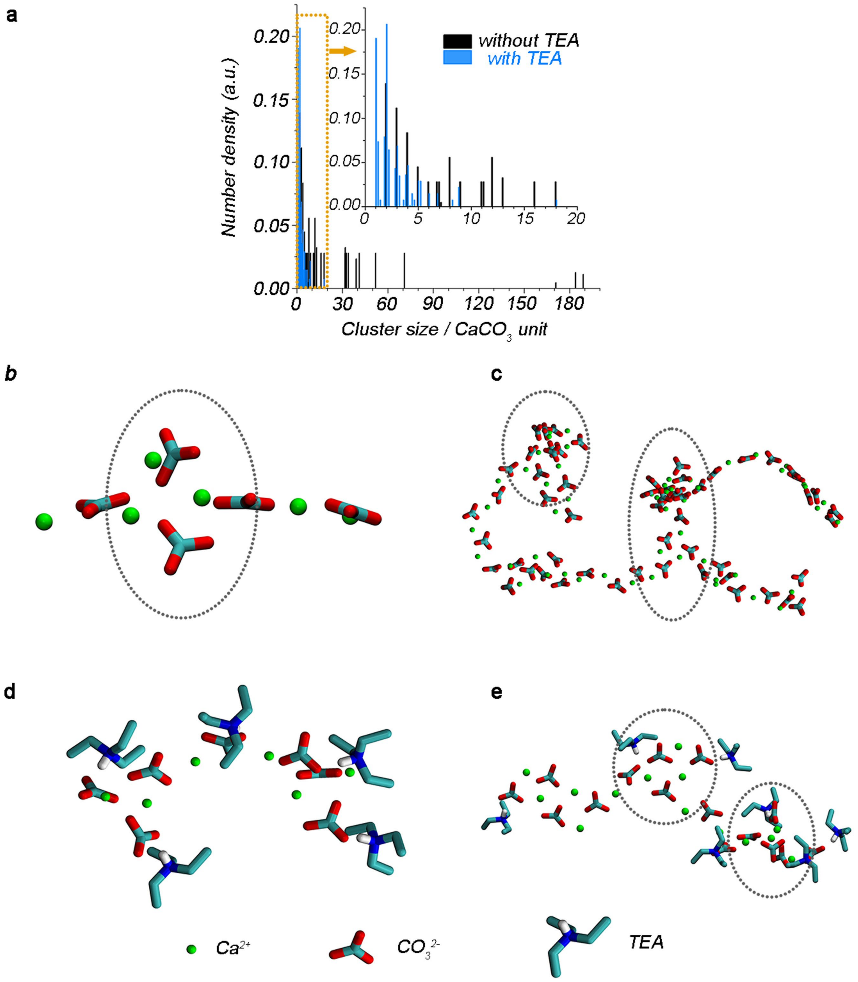
**Extended Data Fig. 1 | ESI-MS analysis of $(CaCO_3)_n$ oligomers.** **a**, Negative-ion mode analysis of $CaCl_2$ and TEA in ethanol solution. **b**, Negative-ion mode analysis of $(CaCO_3)_n$ oligomers after 5 days, indicating that TEA has a long-term stabilization effect. *n* represents the number of $Ca^{2+}$:$CO_3^{2-}$ units in one $(CaCO_3)_n$ oligomer. **c**, **d**, Positive ion mode analysis of $CaCl_2$ and TEA mixed ethanol solution (**c**) and $(CaCO_3)_n$ oligomers (**d**). **e**, ESI tandem MS analysis of the oligomers with *m*/*z* 368, showing only one fragmentation peak at around *m*/*z* 260 with negative charges. Because no TEA could be detected in the MS–MS peak (the intensity below *m*/*z* 200 is zero), this indicates that no TEA was present in the species with *m*/*z* 368.



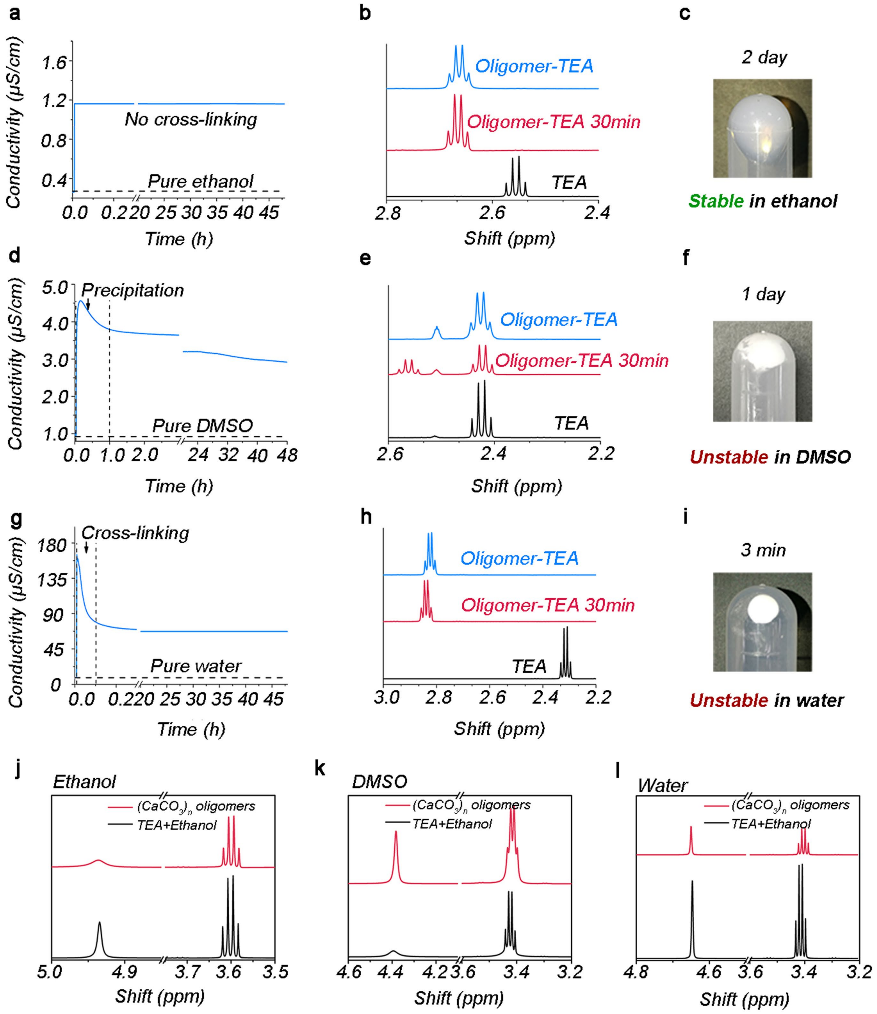
**Extended Data Fig. 2 | Synchrotron SAXS, liquid-cell TEM, dynamic light scattering and viscosity analyses of $(CaCO_3)_n$ oligomers and their aggregates. a**, Scattering plots of different concentrations of $(CaCO_3)_n$ measured by synchrotron SAXS. The red curve is the fitting result from DAMMIF. At a concentration of 57.0 mg $l^{-1}$ an increase in the scattered intensity was detected, and this broad shoulder peak indicated the aggregation of $(CaCO_3)_n$ with a large distribution of sizes. A pronounced maxima of the scattered intensity could be observed at a concentration of 285 mg $l^{-1}$, demonstrating inter-oligomer correlations induced by the formation of larger $(CaCO_3)_n$ aggregates. **b**, Size distribution function of oligomer aggregates at 57.0 mg $l^{-1}$. The inset shows the simulated shape of aggregates, indicating the existence of chain-like structures, including branches at high concentrations. **c**, Liquid-cell TEM of the aggregates of the $(CaCO_3)_n$ oligomers showing a chain-like contrast; the boundary between the background (blue) and sample (yellow) is enhanced in the false colour image on the right, and the chain-like aggregates are marked with arrows. **d**, Viscosity of $(CaCO_3)_n$ oligomers measured at 5 °C ($n \geq 3$). The change in the viscosity with concentration is better fitted with the Huggins equation ($R^2 = 0.98$) than with the Einstein equation ($R^2 = 0.94$), indicating that the shape of the aggregates in solution was more chain-like than spherical. **e**, The hydrodynamic diameter of $(CaCO_3)_n$ oligomers or their aggregates at different oligomer concentrations measured by dynamic light scattering with a Ca:TEA ratio of 1:20. The increase (decrease) of the hydrodynamic diameter is reversible by condensing (diluting) the $(CaCO_3)_n$ oligomer solution. **f**, Reversible aggregations and disaggregations of the $(CaCO_3)_n$ oligomers revealed by synchrotron SAXS. **g**, Scheme of the reversible aggregations and disaggregations of the oligomer unit are controllable by the concentration changes. **h**, In situ liquid-cell TEM observation of $(CaCO_3)_n$ oligomers. The chain aggregates remain in dynamical change (aggregation–disaggregation) states.



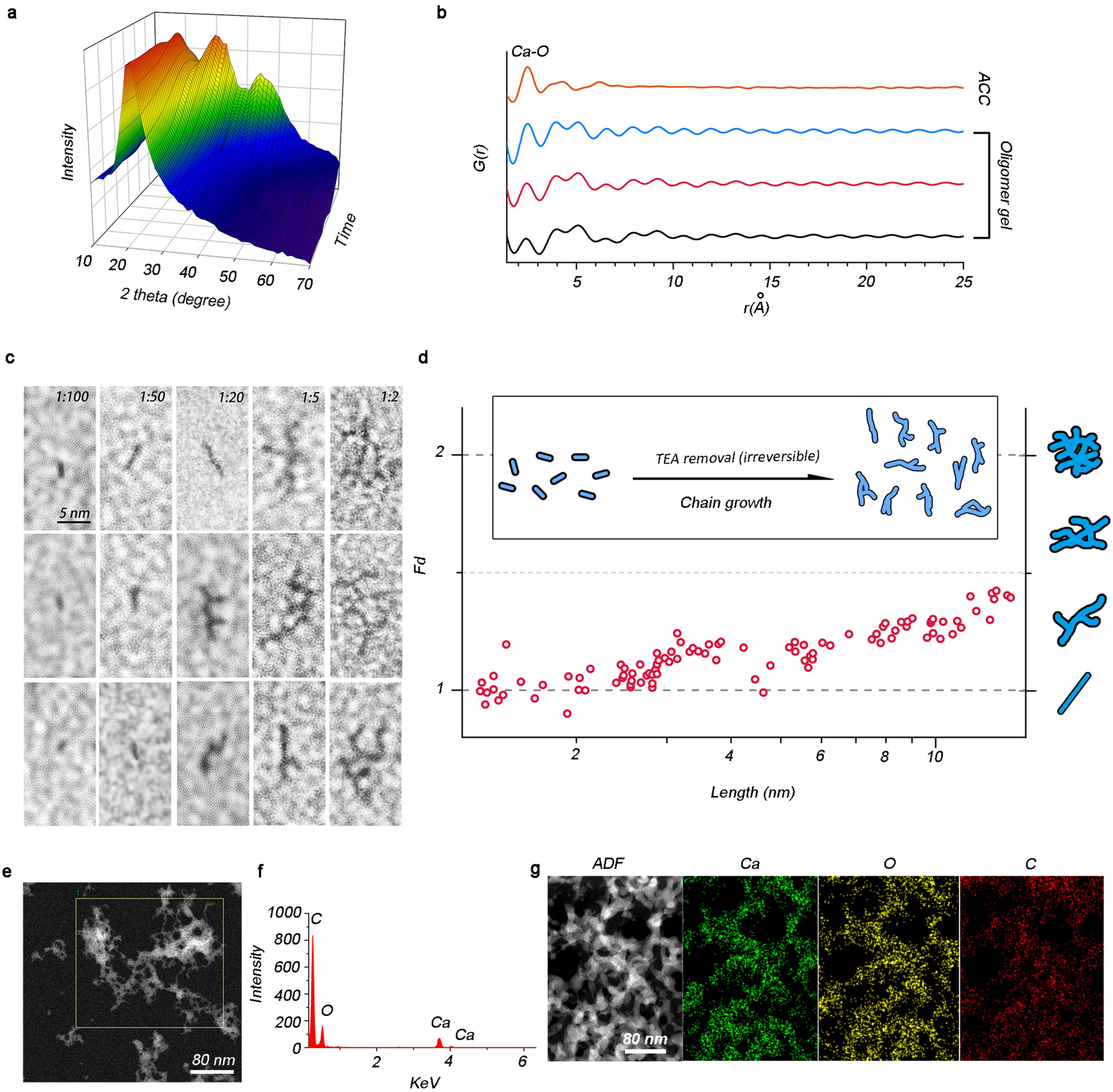
**Extended Data Fig. 3 | FTIR spectra of gel-like $(CaCO_3)_n$ oligomers.** **a**, FTIR spectra showing the peak corresponding to the carbonate group in $(CaCO_3)_n$. **b**, The C–N bond in TEA. **c**, The N···H–O bond between TEA and protonated carbonate[55]. **d**, The spectra between 400 cm$^{-1}$ and 4,000 cm$^{-1}$ are almost the same for ethanol and gel-like oligomers, which confirms that ethanol is the major component in the gel-like $(CaCO_3)_n$ oligomers. **e**, The peak at 3,340 cm$^{-1}$ was contributed by the –OH group of ethanol, which obstructed the signal from H in N···H–O. **f**, The spectrum of ethyl acetate and ethyl acetate-based gel-like $(CaCO_3)_n$ oligomers. The ethanol had been completely removed from the ethyl acetate-based gel-like $(CaCO_3)_n$ oligomers, because the specific peaks of ethanol were not detected; by contrast, the signal at 3,000–4,000 cm$^{-1}$ was detected. **g**, The peak at 3,555 cm$^{-1}$ was contributed by H in the N···H–O bond.



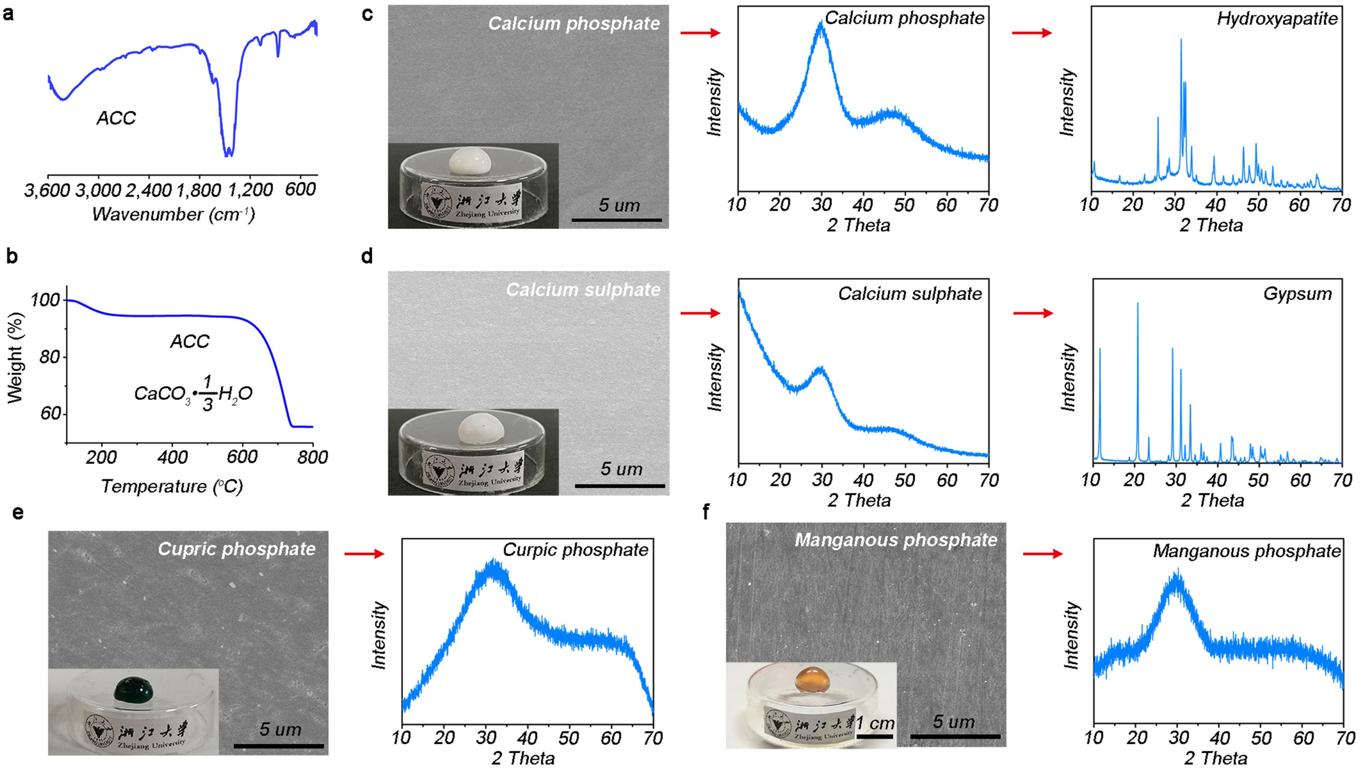
**Extended Data Fig. 4 | Molecular dynamics simulation of the TEA-stabilized $(CaCO_3)_n$.** **a**, Simulated size distribution of $(CaCO_3)_n$ clusters in the presence and absence of TEA. **b**, The branch structure in a $(CaCO_3)_5$ cluster without TEA. The circle shows the branching site. **c**, The branch structure in the large cluster aggregates without TEA. **d**, There is no branch structure in a cluster of $(CaCO_3)_7$ with TEA stabilization. **e**, A branch structure in $(CaCO_3)_{15}$ even with TEA stabilization. These results demonstrate that the stabilization effect of TEA promotes linear growth of $(CaCO_3)_n$.



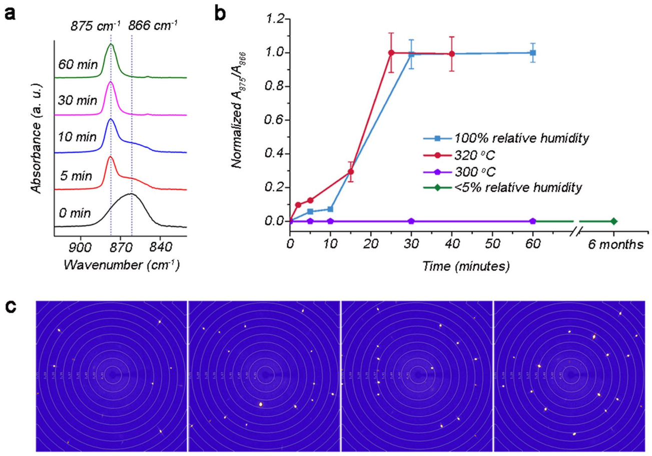
**Extended Data Fig. 5 | $(CaCO_3)_n$ oligomers in ethanol, DMSO and water. a**, **d**, **g**, Conductivity of dispersed oligomer solution with ethanol, DMSO or water as solvent. After high-speed centrifugation and redispersion of $(CaCO_3)_n$ oligomers in ethanol, DMSO or water, the conductivity of the oligomer solution remained steady at 1.2 μS cm$^{-1}$ in ethanol (**a**); the conductivity increased to 4.6 μS cm$^{-1}$ in DMSO and then decreased to 3.8 μS cm$^{-1}$ within 1 h in DMSO (**d**); the conductivity increased to 161.2 μS cm$^{-1}$ but rapidly decreased to 72.0 μS cm$^{-1}$ within 6 min in water (**g**). **b**, **e**, **h**, $^1$H NMR measurement of the ethyl group of TEA (pure or bound with oligomer) in ethanol (**b**), DMSO (**e**) and water (**h**). The results revealed that the chemical shift of $^1$H of the ethyl group of pure TEA was 2.56 ppm in ethanol and 2.67 ppm in the oligomer solution (**b**), and this change in chemical shift can be attributed to the interaction between the TEA and the carbonates in the oligomers. After two days, the unchanged chemical shift of 2.68 ppm corroborated the stable capping effects of the TEA in ethanol. In DMSO (**e**), two signals—2.57 and 2.42 ppm—were attributed to bonded and de-bonded TEA, respectively. In water (**h**), only the peak for de-bonded TEA in the form of [H-TEA]$^+$ could be detected at 2.83 ppm. **c**, **f**, **i**, Photographs of the $(CaCO_3)_n$ materials, which exhibit gel-like features in ethanol and powder-like features in DMSO and water. **j**–**l**, $^1$H NMR measurement of the ethyl group and hydroxyl group of ethanol (pure or bound with oligomer) in ethanol (**j**), DMSO (**k**) and water (**l**). The negligible change in chemical shift proved that no distinct interactions occurred between ethanol and the $(CaCO_3)_n$ oligomers.



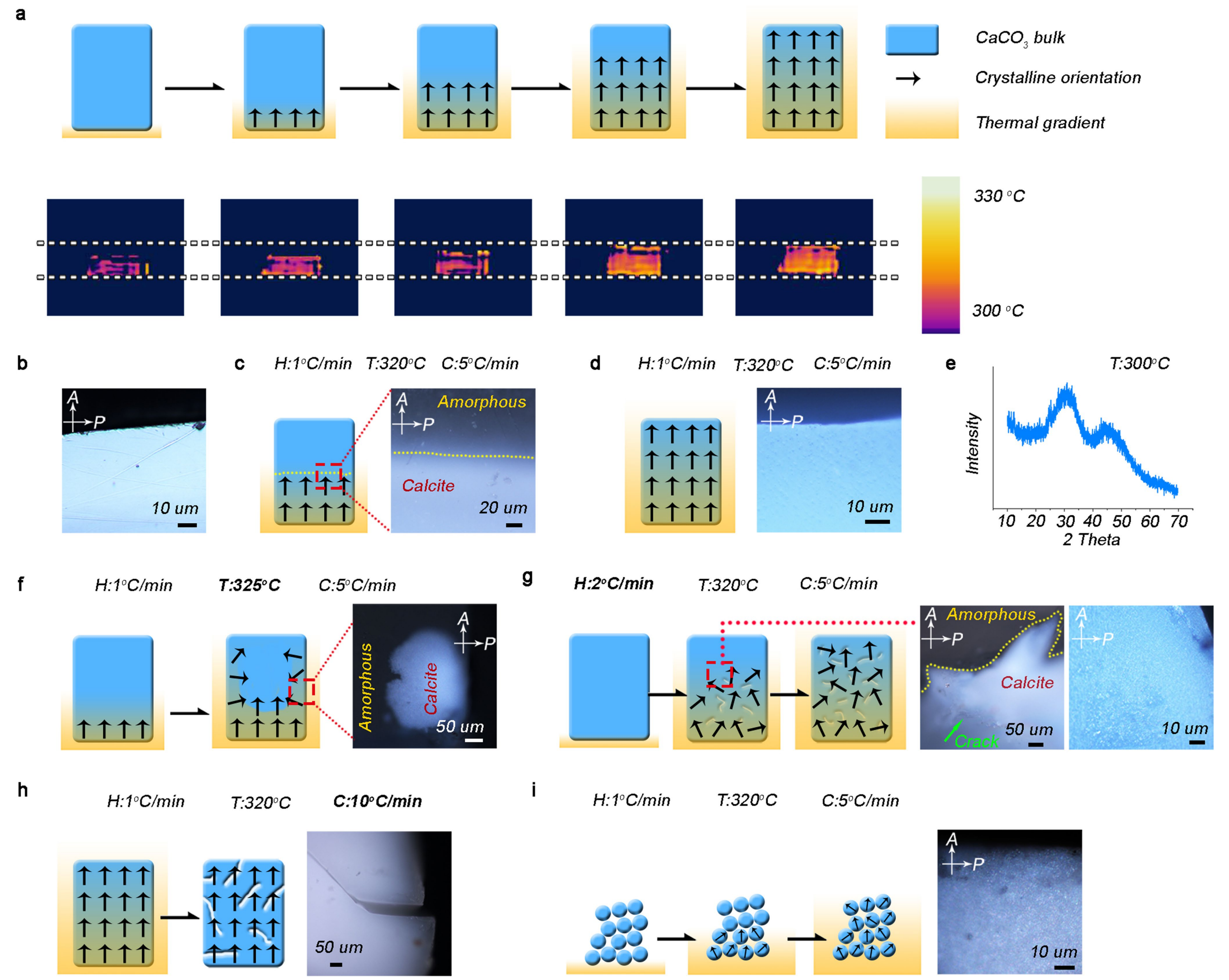
**Extended Data Fig. 6 | Cross-linking of $(CaCO_3)_n$ oligomers. a**, In situ XRD analysis during the drying of $(CaCO_3)_n$ oligomers, showing the evolution from gels to an ACC monolith. **b**, Evolution of the RDF from gels to an ACC monolith. The peak at around 2.4 Å is attributed to Ca–O, and its intensity increased during crosslinking. **c**, Additional high-resolution TEM images of the crosslinking of $(CaCO_3)_n$ upon the removal of TEA, showing the stepwise chain growth and branch formation. **d**, Statistical results of the lengths and fractal dimensions of grown $(CaCO_3)_n$ oligomers. The scheme shows the anisotropic chain growth with branches after TEA removal. **e**, Cross-linked $(CaCO_3)_n$ with a network structure over an ultra-thin carbon support film. **f**, EDS analysis of the crosslinked $(CaCO_3)_n$ in **e**. **g**, EDS mapping analysis of the crosslinked $CaCO_3$ oligomers.



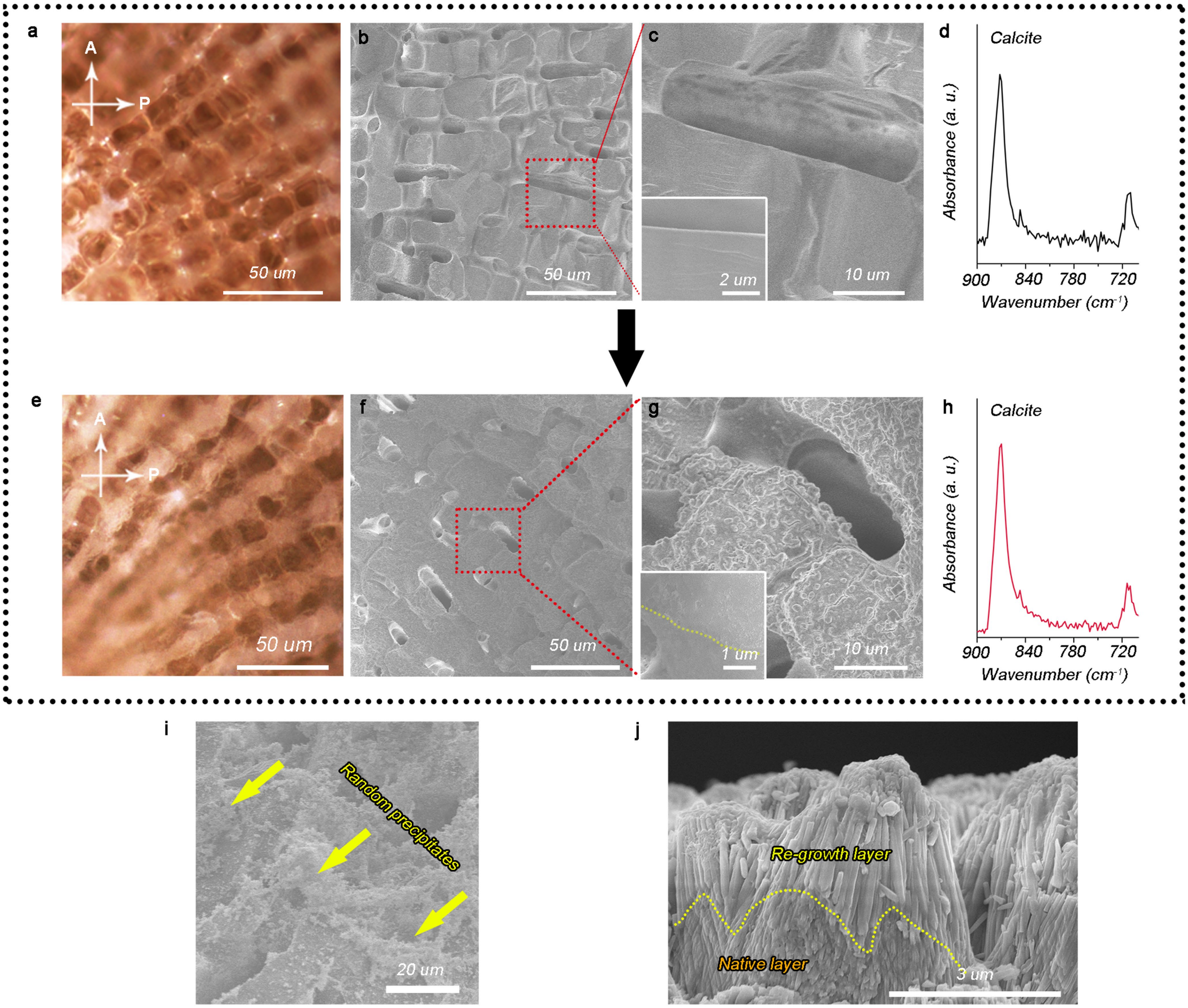
**Extended Data Fig. 7 | Monolithic ACC and other inorganic materials.** **a**, FTIR analysis of the monolithic $CaCO_3$, showing ACC as the only component. **b**, Thermal gravimetric analysis of the monolithic $CaCO_3$, showing a $CaCO_3$:$H_2O$ molar ratio in the ACC of 3:1. **c**–**f**, Photographs and XRD patterns of multiple amorphous monoliths, including calcium phosphate (**c**), calcium sulfate (**d**), cupric phosphate (**e**) and manganous phosphate (**f**) monoliths. The subsequent crystallization manners of these samples are different. The crystallizations of amorphous cupric phosphate and amorphous manganous phosphate cannot be induced by our thermal treatments (700 °C)[56]. Amorphous calcium phosphate and calcium sulfate can transform to hydroxyapatite (**c**) and gypsum (**d**), respectively, by using humidity or water treatments. Although large single crystals without cracks have not been obtained, we believe that future improvements in crystallization control may provide a suitable solution. In conclusion, crosslinking oligomers provides a general strategy for the construction of amorphous monoliths with continuous structures, but an appropriate crystallization treatment is also required to extend the application of this method to the production of crystalline monoliths.



**Extended Data Fig. 8 | Crystallization of ACC under different conditions. a**, FTIR spectra of the monolithic $CaCO_3$ at different time periods under 100% relative humidity at 25 °C. The peaks at 866 cm$^{-1}$ belong to the amorphous phase, and those at 875 cm$^{-1}$ belong to the crystalline phase. The change of the peaks with time indicates a humidity-induced crystallization process. **b**, Kinetics of phase transformation from ACC to calcite under different conditions: <5% relative humidity treatment at 25 °C, 100% relative humidity treatment at 25 °C, 300 °C thermal treatment, and 320 °C thermal treatment. The results show the treatment condition controls for the crystallization. It should be noted that although single crystals can be generated by humidity treatment, the size limitation is stricter than that for single crystals generated by thermal treatment. **c**, X-ray diffraction pattern of thermal induced calcite sample measured along <100>, <010>, <001> and <111> zone axes. The indexing rate of our sample was 79%, while that of geological single-crystal calcite was 87%.



**Extended Data Fig. 9 | Control of thermal-induced crystallization.** **a**, Scheme of thermal-induced single-crystalline calcite formation. The infrared thermal images show the thermal gradients from the proximal heating site to the distal site during the treatment, which is the key to ensure the oriented crystallization for single-crystal formation. **b**, POM image of a single crystalline calcite. **c**–**i**, Thermal treatment at different conditions demonstrating the importance of selecting an appropriate protocol . *H, T* and *C* represent the heating rate, transformation temperature, and cooling rate, respectively. **c**, **d**, The continuous movement of the complete amorphous-crystalline crystallization frontier induced by the well-organized thermal gradient with the recommend protocol of $H = 1\,°\mathrm{C\,min^{-1}}$, $T = 320\,°\mathrm{C}$ and $C = 5\,°\mathrm{C\,min^{-1}}$. **e**, XRD pattern of ACC with a protocol of $T = 300\,°\mathrm{C}$, showing no crystallization at the low heating temperature. **f**, A protocol of $H = 1\,°\mathrm{C\,min^{-1}}$, $T = 325\,°\mathrm{C}$ and $C = 5\,°\mathrm{C\,min^{-1}}$ results in a break of the thermal gradient; a spontaneous calcite crystallization is induced at a distal site rather than the amorphous–crystalline crystallization frontier, resulting in polycrystalline formation. **g**, A protocol of $H = 2\,°\mathrm{C\,min^{-1}}$, $T = 320\,°\mathrm{C}$ and $C = 5\,°\mathrm{C\,min^{-1}}$ also results in a thermal gradient break so that the random crystallizations occur at the amorphous–crystalline interface. **h**, A protocol of $H = 1\,°\mathrm{C\,min^{-1}}$, $T = 320\,°\mathrm{C}$ and $C = 10\,°\mathrm{C\,min^{-1}}$ causes the formation of cracks in the resulting calcite owing to the fast cooling rate. **i**, Using the ACC bulk produced by the conventional particle packing, polycrystalline formation is induced after a thermal treatment of $H = 1\,°\mathrm{C\,min^{-1}}$, $T = 320\,°\mathrm{C}$ and $C = 5\,°\mathrm{C\,min^{-1}}$, which is attributed to its discontinuous internal structure.



**Extended Data Fig. 10 | Repair of sea-urchin spine and enamel by a combination of the crosslinking oligomers. a**, POM image of natural sea-urchin spine. **b**, **c**, SEM images of natural sea-urchin spine, showing the complicated structure with a smooth surface. Inset in **c** is a cross-sectional view of $CaCO_3$ on the spine. **d**, ATR-FTIR characterization of the sea-urchin spine, which is inorganic calcite. **e**, POM image of the sea-urchin spine grown using the $(CaCO_3)_n$ oligomers; the crystallization is induced by a simulated sea water at 25 °C. **f**, **g**, SEM images of the repaired sea-urchin spine, which maintains the original complicated structure. A magnified image (**g**) shows the surface of freshly grown $CaCO_3$. The inset of **g** is a cross-sectional view of the $CaCO_3$ grown on the spine, demonstrating the continuous interfacial structure between the native sea-urchin spine and the grown calcite. **h**, ATR-FTIR characterization of the repaired layer on the sea-urchin spine, which is also calcite. **i**, SEM image of the repaired sea-urchin spine using a conventional solution crystallization method. The random precipitation of numerous calcite particles rather than the expected oriented growth is resulted, demonstrating a failure in the biomineral repair. **j**, Regrowth of enamel structure by using calcium phosphate oligomers and crystallization is induced by simulated oral fluid at 37 °C. Epitaxial growth of the enamel rods is observed, which is due to the continuous construction by the crosslinking oligomers.

LETTER

https://doi.org/10.1038/s41586-019-1638-9

# A large source of cloud condensation nuclei from new particle formation in the tropics

Christina J. Williamson[1,2]*, Agnieszka Kupc[2,3], Duncan Axisa[4,9], Kelsey R. Bilsback[5], ThaoPaul Bui[6], Pedro Campuzano-Jost[1,7], Maximilian Dollner[3], Karl D. Froyd[1,2], Anna L. Hodshire[5], Jose L. Jimenez[1,7], John K. Kodros[5,10], Gan Luo[8], Daniel M. Murphy[2], Benjamin A. Nault[1,7], Eric A. Ray[1,2], Bernadett Weinzierl[3], James C. Wilson[4], Fangqun Yu[8], Pengfei Yu[1,2,11], Jeffrey R. Pierce[5] & Charles A. Brock[2]

**Cloud condensation nuclei (CCN) can affect cloud properties and therefore Earth's radiative balance[1–3]. New particle formation (NPF) from condensable vapours in the free troposphere has been suggested to contribute to CCN, especially in remote, pristine atmospheric regions[4], but direct evidence is sparse, and the magnitude of this contribution is uncertain[5–7]. Here we use in situ aircraft measurements of vertical profiles of aerosol size distributions to present a global-scale survey of NPF occurrence. We observe intense NPF at high altitudes in tropical convective regions over both Pacific and Atlantic oceans. Together with the results of chemical-transport models, our findings indicate that NPF persists at all longitudes as a global-scale band in the tropical upper troposphere, covering about 40 per cent of Earth's surface. Furthermore, we find that this NPF in the tropical upper troposphere is a globally important source of CCN in the lower troposphere, where CCN can affect cloud properties. Our findings suggest that the production of CCN as new particles descend towards the surface is not adequately captured in global models, which tend to underestimate both the magnitude of tropical upper tropospheric NPF and the subsequent growth of new particles to CCN sizes.**

New particles form in the atmosphere when condensing gases form stable clusters with diameters of more than 1.5 nm or so[8]. Growth by condensation and coagulation may enable particles to reach diameters of more than around 60 nm, at which point they can act as CCN. Atmospheric observations are required to guide the incorporation of NPF mechanisms into models[9]. Large numbers of small particles have previously been observed at high altitude in the tropics[10–12], because deep convective clouds loft condensable vapours and remove most larger particles that would otherwise compete with NPF as sinks for these vapours[13] (Fig. 1). Newly formed particles grow to CCN sizes in subsiding air outside of the convective clouds[14].

Global-scale measurements are needed to understand the scale and impact of NPF in the upper troposphere. However, satellites cannot detect particles with diameters of less than 100 nm, and previous in situ observations have been of regional scale[11–13]. To address this, as part of the NASA Atmospheric Tomography Mission (ATom)[15] we conducted in situ, global-scale measurements of particle size distributions over the Pacific and Atlantic oceans during multiple seasons, with near pole-to-pole coverage and systematic profiling between

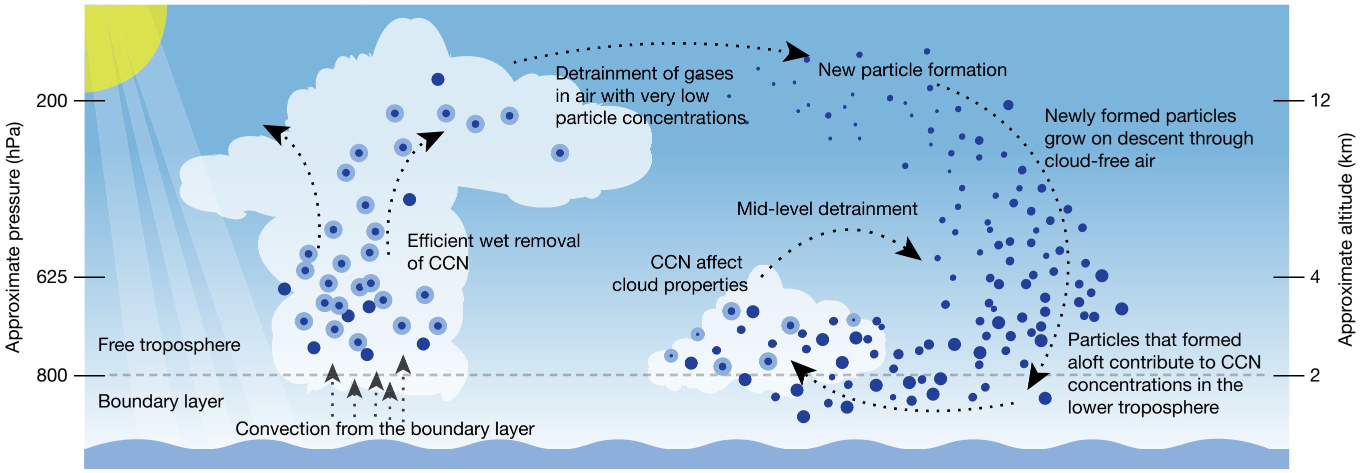


**Fig. 1 | New particle formation and growth to CCN sizes in the tropical convective region.** Left, deep convective clouds loft air from the boundary layer. Larger particles (CCN; dark blue circles) become activated to produce cloud droplets (light blue circles with dark centres) and are removed through precipitation and wet deposition ('wet removal'), reducing the condensation sinks that are found in air being moved from the cloud to the surrounding air ('detrained') at high altitude. Less-soluble aerosol precursors detrain, oxidize and form new particles, which grow by condensation and coagulation as they descend (right), with many reaching CCN sizes before they reach the top of the boundary layer. These CCN duly affect cloud properties. Image courtesy of K. Bogan, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder.

---

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA. [2]Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA. [3]Faculty of Physics, Aerosol Physics and Environmental Physics, University of Vienna, Vienna, Austria. [4]Department of Mechanical and Materials Engineering, University of Denver, Denver, CO, USA. [5]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA. [6]Earth Science Division, NASA Ames Research Center, Moffett Field, CA, USA. [7]Department of Chemistry, University of Colorado, Boulder, CO, USA. [8]Atmospheric Sciences Research Center, State University of New York at Albany, Albany, NY, USA. [9]Present address: Droplet Measurement Technologies, Longmont, CO, USA. [10]Present address: Institute of Chemical Engineering Sciences, ICE/FORTH, Patras, Greece. [11]Present address: Institute for Environment and Climate Research, Jinan University, Guangzhou, China. *e-mail: christina.williamson@noaa.gov

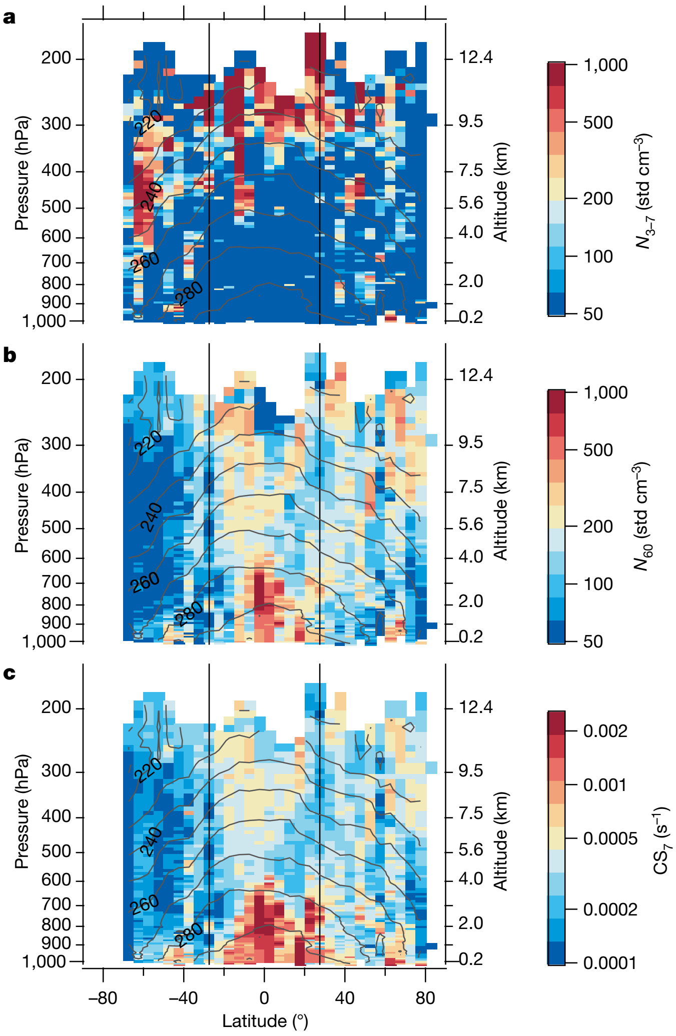


**Fig. 2 | Average aerosol properties from ATom 1 and 2. a**, Number concentrations of particles with diameters of between 3 nm and 7 nm, revealing high concentrations at high altitudes in the tropics. **b**, **c**, Average number concentration of particles bigger than 60 nm (**b**) and the condensation sink from all particles larger than 7 nm ($CS_7$; **c**), revealing mid to low values at high altitudes in the tropics. Grey curves show temperature contours (with 10 K spacing); black vertical lines show the latitude range of the TCRs. std cm$^{-3}$, per cubic centimetre at standard temperature and pressure.

altitudes of roughly 0.18 km and 12 km (Extended Data Fig. 1). We observed evidence for abundant recent NPF at high altitudes within the tropical convective region (TCR; Figs. 2, 3 and Extended Data Fig. 2) and for subsequent growth of the particles as they subsided, and we calculated both the gas-phase condensation sink from all particles with diameters greater than 7 nm ($CS_7$), and the particle coagulation sink, which together govern the probability that particles will form and grow to reach CCN sizes (see Methods). We observed this phenomenon in both August and February, revealing seasonal persistence.

In the TCR, high number concentrations occur over roughly 30° of latitude over both Atlantic and Pacific oceans (Extended Data Fig. 2). NPF has also been observed at high altitude over the Amazon[16]. We compare our observations with the results of four global-scale chemical-transport models that have explicit size-resolved aerosol microphysics (Extended Data Table 1). Some models that reproduce the observed pattern of small particles from NPF in the TCR indicate that these particles persist as a nearly continuous band around the global tropics (Extended Data Fig. 3). Together, these observations and models indicate that tropical NPF covers roughly 40% of Earth's surface.

Deep convective clouds in the tropics enable NPF by removing pre-existing particles that are larger than 60 nm, thus reducing the sinks for small particles and condensable vapours[13,17]. However, our measurements—which allow direct calculation of condensation and coagulation rates—show that tropical convection does not produce sinks with uniquely low $CS_7$ values at high altitude in the TCR, where the most recent NPF is observed (Fig. 2b, c and Extended Data Fig. 4). Therefore, further explanation is required for the consistently high numbers of particles from NPF observed in this region compared with regions with even lower $CS_7$ values.

Low temperatures increase rates of NPF[18]. If condensable vapours were uniform across the atmosphere, we would expect to see the most particles produced by NPF at the lowest available temperatures and at the lowest $CS_7$ values. Globally, NPF occurs at temperatures of less than 270 K and at $CS_7$ values of less than $8 \times 10^{-4}$ s$^{-1}$(Figs. 2, 3). Within the TCR, NPF occurs mainly at the lowest available $CS_7$ and temperature; however, in other regions of the troposphere, there are much lower $CS_7$ values and temperatures but weaker or no NPF. Therefore, a stronger source of condensable material must be available at high altitudes in the TCR than in other cold, low-$CS_7$ areas (except in the Southern Ocean in February). This probably results from a combination of convective activity in this region bringing precursor gases from lower altitudes (Fig. 1), and high solar elevation angles in the tropics increasing the availability of hydroxyl radicals (OH) to produce condensable vapours at faster rates[19].

Within the TCR, median particle diameters increase fairly continuously with decreasing altitude (Fig. 4a). This must indicate particle growth over time, given that these data are from cloud-free air (see Methods), which has a general descending motion[20] and remains mostly in the tropics (Extended Data Fig. 5). The observed increase in particle size with decreasing altitude is inconsistent with mixing from continental boundary layer air, where size distributions instead show a substantial accumulation mode (with particles of roughly 60–500 nm in size) or nucleation mode (roughly 2–12 nm), accompanying the Aitken mode (roughly 12–60 nm) (Extended Data Fig. 6b).

High number concentrations of particles with diameters greater than 60 nm ($N_{60}$; large enough to act as CCN in convective clouds at less than 1% supersaturation[21]) are seen in the TCR (Fig. 4c) (although $N_{60}$ is fairly constant, the fraction of particles that are bigger than 60 nm increases with decreasing altitude). Air at middle and lower altitudes in the TCR is a mix of descending air, which has reduced relative humidity, and air from lower altitudes, with higher relative humidity (Extended Data Fig. 7). However, $N_{60}$ increases with decreasing relative humidity at mid-altitudes (Fig. 4d), indicating that CCN-sized particles are more abundant in the dry descending air than in the moist air from lower altitudes. Particles of 30–60 nm exist in the TCR marine boundary layer (MBL), but are unlikely to be primary particles[22,23]. With no evident source within the tropical MBL (no strong NPF events (Fig. 2a) and few nucleation-mode particles (Extended Data Fig. 6c)), these particles may originate from the free troposphere. Previous studies have detected NPF at the top of the MBL[24,25], but we do not observe this within the TCR (although we did observe it at other latitudes), and so conclude that this does not contribute substantially to particle concentrations here. Therefore, NPF at high altitude in the TCR appears to increase CCN concentrations in the lower troposphere, where these CCN can affect cloud properties and thus the global radiation budget[7,26]. Preliminary modelling studies suggest a global radiative effect of around −0.1 W m$^{-2}$ from this CCN source (Extended Data Fig. 9).

Models show abundant NPF at high altitude in the TCR (although fewer nucleation-mode particles are produced than in the observations). But, except for the CAM5-APM model (which greatly overestimates gas-phase volatile organic compounds at high altitude[27]), these models produce much lower concentrations of CCN-sized particles than we observed (Fig. 4b, c).

High-altitude TCR NPF occurs at higher $CS_7$ values in our observations than in all models except CAM5-APM (Fig. 3d). Models also produce fewer small particles at high altitude than we observed (Fig. 4a, b and Extended Data Fig. 6a). The overall performance of these models

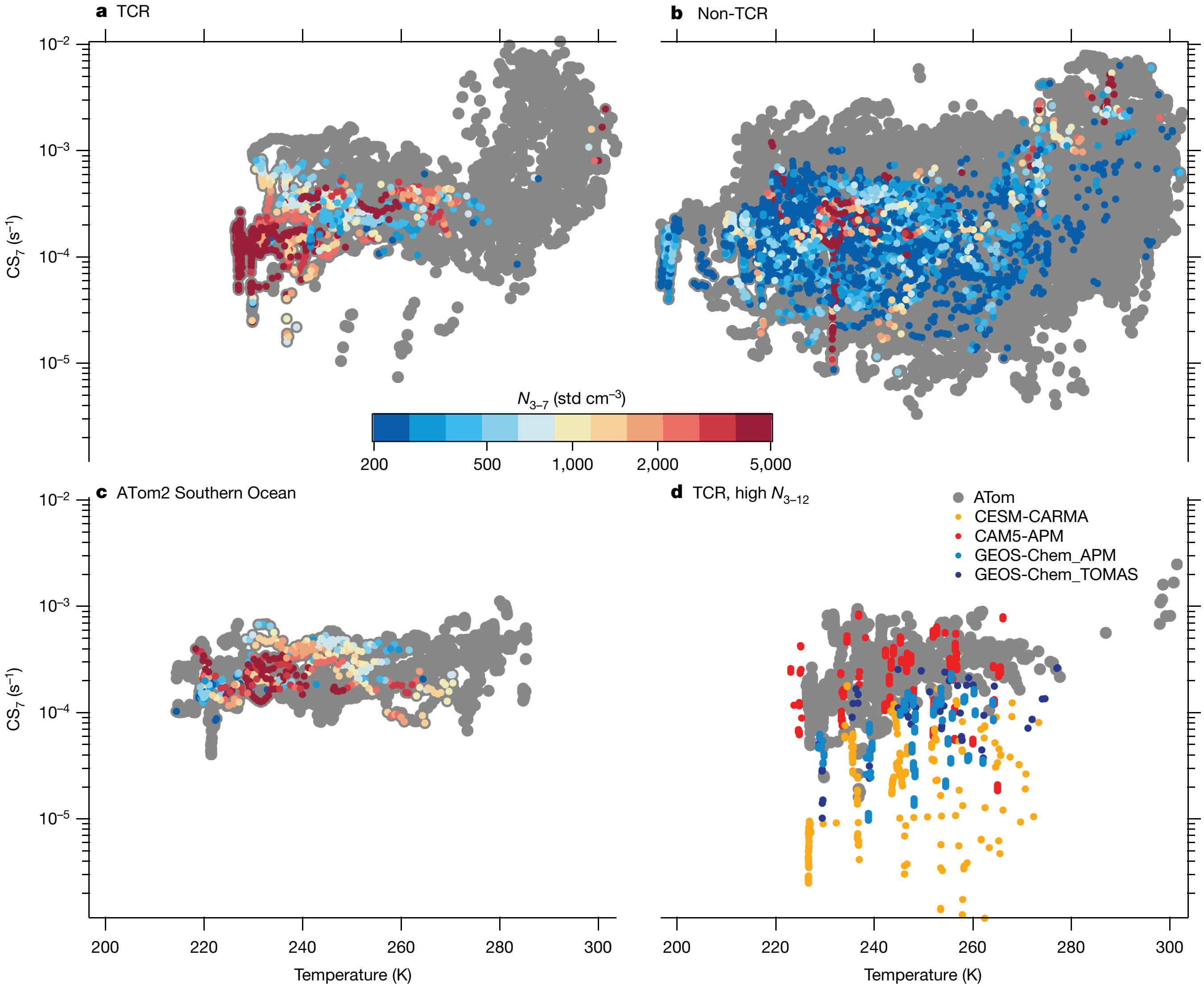


**Fig. 3 | The relationship between $CS_7$, temperature and NPF. a–c,** Temperature and $CS_7$ values for the TCR (**a**), non-TCR (**b**) and Southern Ocean in February (c), with colours showing statistically significant concentrations of particles with diameters of 3–7 nm ($N_{3-7}$) (see Methods for definition of 'statistically significant concentrations'). Grey points indicate data for which concentrations of $N_{3-7}$ were not statistically significant. Higher concentrations of 3–7-nm particles are observed for given temperatures and $CS_7$ values within the TCR than elsewhere (except in the Southern Ocean during February), suggesting a greater availability of condensable vapours in the TCR. **d**, Observed (grey) and modelled (coloured) $CS_7$ and temperature values in the TCR for number concentrations of 3–12-nm particles that exceed 1,750 $cm^{-3}$. All models except CAM5-APM show high concentrations of 3–12-nm particles at lower $CS_7$ values than observed.

suggests a deficit of condensable material and/or missing nucleation or growth mechanisms, especially because increasing $CS_7$ values to match the observations would reduce the modelled particle concentrations even further. Previous observations[11] also underestimated sinks in this region by considering only particles that are larger than 60 nm, so the higher observed $CS_7$ values from ATom (Figs. 2b, 3a) represent a departure from the models and from previous observations.

Sensitivity analysis (see Methods and Extended Data Fig. 9) shows that errors in nucleation rates or mechanisms are unlikely to cause the underprediction of particle concentrations in the models, given that substantial scaling of nucleation rates did not produce a substantial change in the particle size distribution and resulting number of CCN. This is due to a feedback mechanism, whereby increasing nucleation rates slows growth rates and increases coagulation rates, thus dampening the sensitivity of CCN to changes in nucleation[28]. We also find that uncertainty in the amount of inorganic condensable material has little effect on the resulting number of CCN.

Missing organics could also explain the underprediction: none of the models used here includes organic-mediated nucleation. Organics in the free troposphere are well known to be a dominant contributor to nanoparticle growth in the boundary layer and under many conditions[29], although they remain poorly simulated in many models[30]. Organic matter contributes substantially to the mass of 50–500-nm particles at high altitudes in the TCR (Extended Data Fig. 7). All models except CAM5-APM underestimate organic mass at high altitudes, suggesting that they are missing the organic species or growth mechanisms necessary to reproduce these observations.

The underprediction of $N_{60}$ particles at lower altitudes suggests that models may scavenge growing particles too efficiently during their descent. Convective wet scavenging and aqueous processing in the GEOS-Chem-TOMAS, GEOS-Chem-APM and CAM5-APM models affect all particles in a grid box, as particles in cloudy and clear portions mix at each time-step, but in reality the cloudy and clear regions do not tend to mix quickly. Aqueous sulfate formation in activated aerosol in convective clouds causes a bimodal structure. This is observed at higher altitude in the modelled size distributions than in the observations (Extended Data Fig. 6), perhaps because of excessive removal and processing of clear-sky aerosol by convective clouds in the models[31].

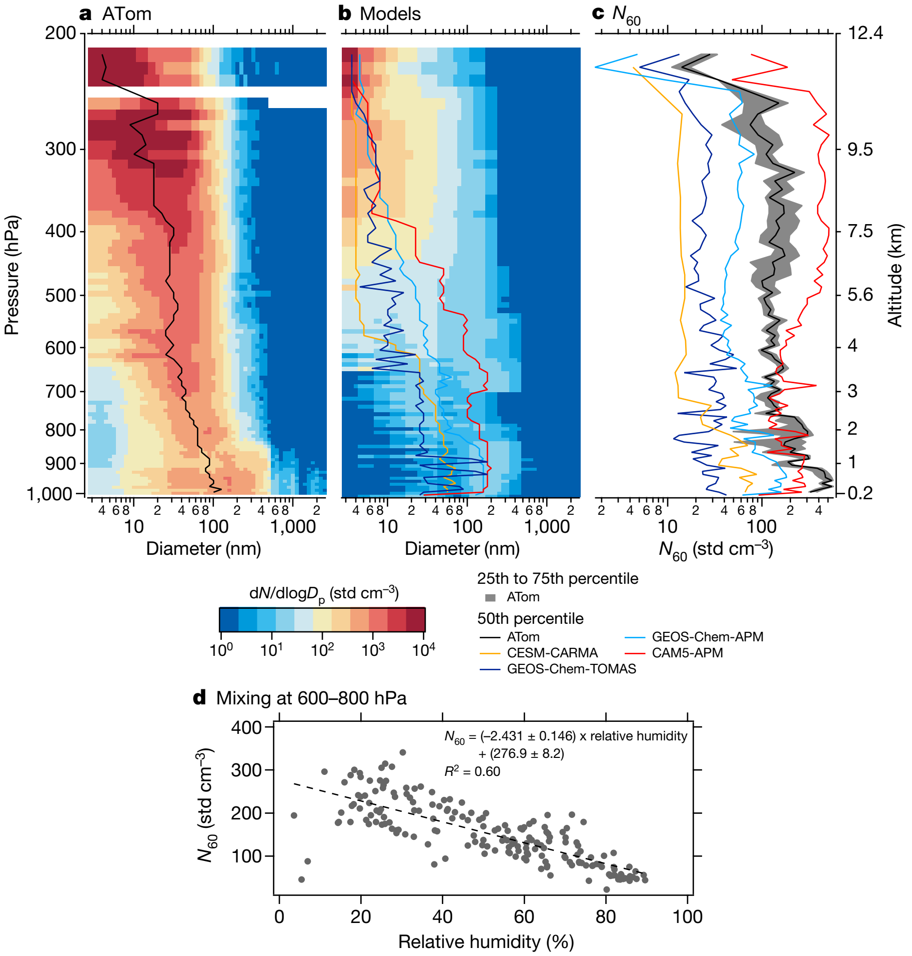


**Fig. 4 | Evidence for particle growth on descent.** All data are from the TCR. **a**, **b**, Average number size distribution as a function of pressure (having removed geographical regions with biomass burning and dust plumes), for ATom observations (**a**) and combined models (**b**). Mode diameters are shown with solid lines. $D_p$, particle diameter. **c**, Observed and modelled number concentrations of particles with diameters greater than 60 nm ($N_{60}$). **d**, Observed $N_{60}$ values (grey data points) plotted against relative humidity at pressures of between 600 and 800 hPa, with a linear least-squares regression fit (dashed line). A relative humidity of more than 10% indicates mixing between descending and lower-altitude air. The increase in $N_{60}$ values with decreasing relative humidity shows that descending air contains more CCN-sized particles than lower-altitude air.

Reducing cloud-processing on descent (a proxy for correcting the clear-sky removal of aerosol) allows modelled $N_{60}$ and $CS_7$ values to better represent the ATom observations (see Methods and Extended Data Fig. 9). A full correction would need to track in-cloud particles between simulation time-steps. The CESM-CARMA model, which correctly accounts for sub-grid clear-sky and cloudy aerosol tracking, shows a larger increase in $N_{60}$ values with altitude than do the other models (Fig. 4), indicating that this correction may allow models with more complex nucleation schemes to capture the resulting increase in $N_{60}$ observed in ATom.

It is therefore likely that many models underestimate CCN in the remote tropical lower troposphere because, through physically incorrect sub-grid aqueous aerosol processing and removal, they remove too many of the particles that form in the tropical upper troposphere. Additional factors might be missing organics, or missing mechanisms for including them in nucleation and growth. This is important for estimating aerosol-cloud radiation effects, as the radiative effect of NPF in the tropical upper troposphere could be of the order of 0.1 W m$^{-2}$ globally (see Methods and Extended Data Fig. 9).

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1638-9.

Received: 4 February 2019; Accepted: 16 July 2019;
Published online 16 October 2019.

1. Rosenfeld, D., Sherwood, S., Wood, R. & Donner, L. Climate effects of aerosol-cloud interactions. *Science* **343**, 379–380 (2014).
2. Boucher, O. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 7 (Cambridge Univ. Press, 2013).
3. Dusek, U. et al. Size matters more than chemistry for cloud-nucleating ability of aerosol particles. *Science* **312**, 1375–1378 (2006).
4. Gordon, H. et al. Causes and importance of new particle formation in the present-day and preindustrial atmospheres. *J. Geophys. Res. Atmos.* **122**, 8739–8760 (2017).
5. Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J. & Carslaw, K. S. Impact of nucleation on global CCN. *Atmos. Chem. Phys.* **9**, 8601–8616 (2009).
6. Pierce, J. R. & Adams, P. J. Uncertainty in global CCN concentrations from uncertain aerosol nucleation and primary emission rates. *Atmos. Chem. Phys.* **9**, 1339–1356 (2009).
7. Clarke, A. D. et al. Free troposphere as a major source of CCN for the equatorial pacific boundary layer: long-range transport and teleconnections. *Atmos. Chem. Phys.* **13**, 7511–7529 (2013).
8. Kulmala, M. et al. Direct observations of atmospheric aerosol nucleation. *Science* **339**, 943–946 (2013).
9. Dunne, E. M. et al. Global atmospheric particle formation from CERN CLOUD measurements. *Science* **354**, 1119–1124 (2016).
10. Brock, C. A., Hamill, P., Wilson, J. C., Jonsson, H. H. & Chan, K. R. Particle formation in the upper tropical troposphere—a source of nuclei for the stratospheric aerosol. *Science* **270**, 1650–1653 (1995).
11. Clarke, A. D. et al. Nucleation in the equatorial free troposphere: favorable environments during PEM-Tropics. *J. Geophys. Res. Atmos.* **104**, 5735–5744 (1999).

12. Weigel, R. et al. In situ observations of new particle formation in the tropical upper troposphere: the role of clouds and the nucleation mechanism. *Atmos. Chem. Phys*. **11**, 9983–10010 (2011).
13. Clarke, A. D. et al. Particle production in the remote marine atmosphere: cloud outflow and subsidence during ACE 1. *J. Geophys. Res. Atmos*. **103**, 16397–16409 (1998).
14. Clarke, A. D. & Kapustin, V. N. A Pacific aerosol survey. Part I: a decade of data on particle production, transport, evolution, and mixing in the troposphere. *J. Atmos. Sci*. **59**, 363–382 (2002).
15. Wofsy, S. C. et al. *ATom: merged atmospheric chemistry, trace gases, and aerosols*. https://daac.ornl.gov/ATOM/guides/ATom_merge.html (ORNL DAAC, 2018).
16. Andreae, M. O. et al. Aerosol characteristics and particle production in the upper troposphere over the Amazon Basin. *Atmos. Chem. Phys*. **18**, 921–961 (2018).
17. Ekman, A. M. L., Wang, C., Strom, J. & Krejci, R. Explicit simulation of aerosol physics in a cloud-resolving model: aerosol transport and processing in the free troposphere. *J. Atmos. Sci*. **63**, 682–696 (2006).
18. Yue, G. K. & Hamill, P. Homogeneous nucleation rates of $H_2SO_4$-$H_2O$ aerosol particles in air. *J. Aerosol Sci*. **10**, 609–614 (1979).
19. Gao, R. S. et al. OH in the tropical upper troposphere and its relationships to solar radiation and reactive nitrogen. *J. Atmos. Chem*. **71**, 55–64 (2014).
20. Folkins, I. & Martin, R. V. The vertical structure of tropical convection and its impact on the budgets of water vapor and ozone. *J. Atmos. Sci*. **62**, 1560–1573 (2005).
21. Fan, J. W. et al. Substantial convection and precipitation enhancements by ultrafine aerosol particles. *Science* **359**, 411–418 (2018).
22. Raes, F. Entrainment of free tropospheric aerosols as a regulating mechanism for cloud condensation nuclei in the remote marine boundary layer. *J. Geophys. Res. Atmos*. **100**, 2893–2903 (1995).
23. Quinn, P. K., Coffman, D. J., Johnson, J. E., Upchurch, L. M. & Bates, T. S. Small fraction of marine cloud condensation nuclei made up of sea spray aerosol. *Nat. Geosci*. **10**, 674–679 (2017).
24. Hegg, D. A., Radke, L. F. & Hobbs, P. V. Particle production associated with marine clouds. *J. Geophys. Res. Atmos*. **95**, 13917–13926 (1990).
25. Hoppel, W. A., Frick, G. M., Fitzgerald, J. & Larson, R. E. Marine boundary layer measurements of new particle formation and the effects nonprecipitating clouds have on aerosol-size distribution. *J. Geophys. Res. Atmos*. **99**, 14443–14459 (1994).
26. Kazil, J. et al. Modeling chemical and aerosol processes in the transition from closed to open cells during VOCALS-REx. *Atmos. Chem. Phys*. **11**, 7491–7514 (2011).
27. Yu, F. A secondary organic aerosol formation model considering successive oxidation aging and kinetic condensation of organic compounds: global scale implications. *Atmos. Chem. Phys*. **11**, 1083–1099 (2011).
28. Raes, F., VanDingenen, R., Cuevas, E., VanVelthoven, P. F. J. & Prospero, J. M. Observations of aerosols in the free troposphere and marine boundary layer of the subtropical Northeast Atlantic: discussion of processes determining their size distribution. *J. Geophys. Res. Atmos*. **102**, 21315–21328 (1997).
29. Trostl, J. et al. The role of low-volatility organic compounds in initial particle growth in the atmosphere. *Nature* **533**, 527–531 (2016).
30. Tsigaridis, K. et al. The AeroCom evaluation and intercomparison of organic aerosol in global models. *Atmos. Chem. Phys*. **14**, 10845–10895 (2014).
31. Hoppel, W. A., Frick, G. M. & Larson, R. E. Effect of nonprecipitating clouds on the aerosol size distribution in the marine boundary layer. *Geophys. Res. Lett*. **13**, 125–128 (1986).

# METHODS

**Overview of ATom.** Measurements were taken during the first two deployments of NASA's Atmospheric Tomography Mission (ATom; 29 July to 23 August 2016, and 26 January to 21 February 2017), referred to as ATom 1 and 2 respectively. This mission comprised four sets of contiguous flights over the Pacific and Atlantic Ocean basins at latitudes from 81° N to 65° S. The flights focused on the remote marine atmosphere, constantly profiling at altitudes between about 0.18 km and 11–13 km so as to resolve the vertical structure of the atmosphere (Extended Data Fig. 1).

**Summary of aerosol measurements.** Aerosol dry size distributions were measured using a suite of instruments at particle diameters from 2.7 nm to 4.8 μm and at a time resolution of 1 Hz (ref. [32]). Inside the DC-8, a nucleation-mode aerosol size spectrometer (NMASS[33–35])—a custom-built battery of five condensation particle counters, each operated with different 50% detection diameters ($d_{50}$)—provided five channels for particles with diameters of 2.7 nm to 60 nm on ATom 1. Two NMASSes were operated on ATom 2, providing ten distinct channels over the same size range. A commercial optical particle counter (OPC), the ultra-high-sensitivity aerosol spectrometer (UHSAS; Droplet Measurement Technologies), specifically adapted to operate over rapidly changing pressures[36,37], measured particles from 60 nm to 500 nm. A second commercial OPC, the laser aerosol spectrometer (LAS; Thermo-Systems Engineering), extended this distribution to 4.8 μm, which was the upper size limit at which particles could be efficiently sampled.

Size distributions of dried samples (with relative humidities of less than 40%) were measured at 1-Hz resolution, which, given average aircraft ascent/descent rates of roughly 7 m s$^{-1}$, provides roughly 7-m vertical resolution in the atmosphere (with a range from around 11 m at the lowest altitudes to around 5 m at the highest). Fourier transform analysis of concentration measurements from the NMASS revealed statistical noise at frequencies of more than 0.1 Hz (ref. [38]; Extended Data Fig. 8); therefore, we average the data to 0.1 Hz, giving an average vertical resolution of 70 m (corresponding to 8 hPa at an altitude of 1 km and 3 hPa at 10 km).

Cloud particle impaction on aircraft and inlet surfaces generates artefact particles that are then sampled by aerosol instruments[39]. Therefore, we removed all in-cloud size-distribution data from the data set before analysis. Clouds are identified by coarse mode-number concentration (measured with a second-generation cloud, aerosol and precipitation spectrometer, or CAPS; Droplet Measurement Technologies), temperature and relative humidity. Water, ice and mixed-phase clouds are removed.

All concentrations are given at standard temperature and pressure (STP): 1,013 hPa and 0 °C.

**Identifying new particle formation.** To identify recent NPF, we sought times when the concentration in the smallest size channel of an NMASS ($d_{50} = 2.7$ nm) was significantly larger than that in the next-largest channel ($d_{50} = 6.9$ nm). ('Significantly larger' means that the difference in concentrations was greater than could be expected from statistical variations in the sample; Extended Data Fig. 8.)

To identify statistically significant counts in the first channel, we assumed that the counts in each channel could be described by Poisson statistics, such that the variance in the number of counts, $C_i$, in a given channel, i, is given by:

$$\mathrm{Var}(C_i) = C_i$$

The concentration measured in channel i, $N_i$, is:

$$N_i = b_i \times C_i$$

where $b_i$ is a factor calculated from the flow rate, pressure and temperature corrections to STP and live-time correction.

Given that variance follows the relationship

$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$$

we can describe the variance in the concentration in channel i as:

$$\mathrm{Var}(N_i) = {b_i}^2\mathrm{Var}(C_i) = {b_i}^2 C_i$$

Because variance also follows the relationship

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

the standard deviation of the difference between concentrations in channels 1 and 2 in the NMASS is:

$$\sigma_{\mathrm{diff}} = \sqrt{({b_1}^2 C_1 + {b_2}^2 C_2)}$$

We consider that the difference between channel 1 and 2 is significant when:

$$N_1 - N_2 > 3\sigma_{\mathrm{diff}} \qquad (1)$$

For a 5° latitude by 10-hPa box for each ocean basin, we used Equation (1) to calculate the proportion of data points collected (at 0.1-Hz frequency) that show significant concentrations of small particles.

**Sink calculations.** For each point in flight, the condensation kernel for a sulfuric acid molecule with particles of each diameter in the size distribution, at the ambient temperature and pressure, is calculated using the Fuchs expression for the coagulation-rate coefficient[40], substituting a sulfuric acid molecule for one of the particles in order to obtain a condensation sink instead of a coagulation sink (and noting that the accommodation coefficient, $\alpha$, in equation (12) of ref. [40] should read $1/\alpha$). We calculate the diameter of a sulfuric acid molecule from bulk properties according to ref. [41], neglecting temperature effects on the probability-distribution function of monomers, dimers and trimers. Over the measurement conditions, this gives a diameter of between 0.545 nm and 0.552 nm. We take the mass of the sulfuric acid molecule to be 98.079 g mol$^{-1}$ (ref. [42]). We assume each particle to have the density of water (1 kg m$^{-3}$).

This condensation kernel is then multiplied by the number of particles in that size bin and summed over all bins to produce the condensation sink. Coagulation sinks are calculated in the same manner.

The condensation sink for particles larger than 7 nm is correlated with coagulation sinks for particles of various sizes and the total surface areas (Extended Data Fig. 4). Therefore, it is reasonable to assume that the relationships explored herein between NPF, condensation sinks and temperatures also hold for coagulation sinks.

Nucleation and growth rates are not sufficiently constrained by our data to enable us to calculate survival probabilities (that is, the probability of a particle of a given size surviving to reach CCN sizes; these probabilities typically range from around 0.1 to 1, depending on the environment[43]). However, these calculated sinks allow us to place our observations within the context of other studies of NPF and growth to CCN sizes. The condensation sink rate constant in the tropics at pressures higher than 700 hPa is generally below 0.0005 s$^{-1}$ (Fig. 2c). The condensation rate is effectively the coagulation rate of a 'particle' the size of a molecule. The coagulation rate scales with $1/D_{\mathrm{p}}$ of the smaller molecule/particle[40], where $D_{\mathrm{p}}$ is the particle diameter; so the coagulation loss rate for a 3-nm particle here would be 0.0001 s$^{-1}$. Coagulation rates tend to decrease with particle size, and so will be smaller as particles grow through larger sizes and into the accumulation mode. Sources of condensable vapours may be lower in the remote upper troposphere than in the continental boundary layer, so we expect growth rates to be slower here, but condensation and coagulation sinks will also be much lower, because fewer sources exist and deep convective clouds remove many of the larger particles, so loss rates are also lower than expected in the continental boundary layer.

**Back trajectories.** We calculated ten-day back trajectories using the Bowman trajectory model[44] and National Centers for Environmental Prediction (NCEP) global forecast system (GFS) meteorology[45]. Trajectories were initialized each minute along all of the ATom flight tracks. A cluster of 240 trajectories offset slightly both horizontally and vertically from the flight-track locations were initialized each minute as an estimate of the trajectory uncertainties. The cloud fractions were based on satellite-derived clouds[46] within 5° longitude of the flight tracks, averaged within 5° latitude bins.

**Particle composition.** The particle analysis by laser mass spectrometry (PALMS) instrument measures the size and chemical composition of individual aerosol particles with diameters of 150–4,000 nm (ref. [47]). Mass spectral signatures differentiate each particle into a compositional class such as biomass burning, mineral dust, sea salt, sulfate/organic/nitrate mixtures, and others. The number fraction of each particle class averaged over one to three minutes is indicative of relative abundance. Biomass-burning fractions measure the influence of smoke on the aerosol population. Likewise, mineral dust fractions are used to identify dust plumes such as those from the Saharan desert. Where indicated in the text or figures, we excluded data for which biomass burning or dust particle types accounted for more than 40% or 10%, respectively, of total particle number concentration. This was in order to exclude plumes from the analyses. For analysis of organic and sulphate mass, we restricted the PALMS size range to particles of 500 nm or less.

A highly customized, high-resolution time-of-flight aerosol mass spectrometer (HR-ToF-AMS; Aerodyne Research)[48–50] measured non-refractory submicrometre aerosol mass composition at 1-Hz resolution (for particles with diameters of 50–500 nm, the measurement efficiency was 100%; for diameters decreasing to 20 nm and increasing to 700 nm, the counting efficiency decreased to 0). Particles were sampled in situ through a dedicated inlet (a HIMIL inlet[51]) and aerodynamic lens into a vacuum chamber, flash vaporized at 600 °C and analysed by electron-impact time-of-flight mass spectrometry. Overall instrument sensitivity was calibrated every flight day, and sulfate relative ionization efficiencies and instrument particle transmission were calibrated at regular intervals during the missions. For improved sensitivity, the raw mass spectra were averaged and analysed at 46-s intervals (about 300-m vertical resolution). Detection limits (as established by periodic blanks) for organic aerosol and sulfate at that time resolution were on average 75 ng m$^{-3}$ and 10 ng m$^{-3}$, respectively, at STP, and improve by the square root of the number of data points on further averaging.

**Descriptions of models.** We compared the ATom data with the results of four chemical-transport models, namely GEOS-Chem (with aerosol microphysics from either the TOMAS[52] or the APM[27,53] package), CAM5 (with aerosol microphysics from APM; described below); and CESM (with aerosol microphysics from CARMA[54–56]). We matched the location and time of the model outputs to our aircraft measurements. Details are given in Extended Data Table 1.

CAM5-APM incorporated the APM sectional aerosol microphysics with MOZART online chemistry in CAM-Chem[57] following refs. [53,58]. Anthropogenic and biogenic emissions of carbon monoxide and non-methane volatile organic compounds come from the POET (precursors of ozone and their effects in the troposphere) database for 2000 (ref. [59]). The anthropogenic emissions of nitrogen oxides, sulfur dioxide, ammonia, black carbon and organic carbon used in CAM-Chem[57] are replaced by the Harvard–NASA emissions component (HEMCO)[60] for simulation years. Open-fire emissions of black carbon and organic carbon are produced from the International Panel on Climate Change gridded decadal monthly mean forest-fire and grass-fire emissions by the National Center for Atmospheric Research[61]. Sea-salt emissions are based on the size-resolved sea-spray emission scheme developed in ref. [62]. Dust emissions are based on the parameterization developed in ref. [63]. Both sea-salt and dust emissions are calculated using online meteorology simulated by CAM. Thermodynamic equilibrium and aqueous phase chemistry of sulfate/nitrate/ammonium, size-resolved aerosol dry and wet depositions, and aerosol–cloud interactions are considered in the model.

For analyses in which biomass burning and dust plumes were filtered from the ATom data, they were likewise filtered from the model outputs as follows. For CESM-CARMA, a separate model run was completed with biomass burning switched off, and times during which the total mass of dust aerosol exceeded $5 \times 10^{-13}\ \mathrm{kg\ m^{-3}}$ (at STP) were removed. For CAM5-APM and GEOS-Chem-APM, times during which the number fraction of dust or black-carbon particles exceeded 10% and 40%, respectively, were removed. For GEOS-Chem-TOMAS, times during which the number fraction of dust and elemental carbon submicrometre particles exceeded 0.5% were removed.

**Sensitivity studies.** To investigate the causes of model underprediction of CCN from tropical upper tropospheric NPF, we performed model sensitivity studies. We excluded CAM5-APM from this analysis because it is the only model used here in which aerosols are fully coupled to convective clouds and precipitation (which increases run-to-run variability, making it difficult to assess the impact of prescribed changes to the model). In the first study, the nucleation rate at latitudes of between 28° N and 28° S and at altitudes above a pressure of 600 hPa was increased by a factor of ten. In the second study, oceanic emissions of dimethyl sulfide (DMS) were increased by a factor of three globally. DMS contributes to atmospheric sulfur dioxide and is thus a source of condensable inorganic material for forming and growing particles. In the third study, rainout and washout in cloud anvils and large-scale cloud systems, and aqueous oxidation of sulfur dioxide, were reduced by a factor of ten at all altitudes between 28° N and 28° S. This is a proxy for reducing cloud processing of particles and gases in descending air, and indicates the potential effect of reducing the physically incorrect representation of sub-grid cloud processing discussed in the main text. We did not apply this scaling of cloud processing to CESM-CARMA, which is not affected by sub-grid cloud overprocessing.

**Aerosol indirect effects.** We used GEOS-Chem with TOMAS and APM to estimate the magnitude of the aerosol first indirect effect resulting from tropical upper tropospheric NPF[64]. For August 2016 (ATom 1), in the base case and in the case with reduced cloud processing of descending air in the tropics, nucleation was switched off at pressures of less than 600 hPa between 28° N and 28° S. This caused a change in the magnitude of the globally averaged aerosol first indirect effect (Extended Data Fig. 9) of $-0.12$ and $-0.14\ \mathrm{W\ m^{-2}}$ for GEOS-Chem-TOMAS, and $-0.039$ and $-0.031\ \mathrm{W\ m^{-2}}$ for GEOS-Chem-APM, for the base and reduced cloud-processing cases, respectively. GEOS-Chem-APM shows a positive radiative forcing in regions in which liquid-water content in the upper troposphere is elevated. A high liquid-water content indicates stronger vertical transport, and so more particles will be lofted from the lower troposphere in these regions. Turning off nucleation here therefore reduces the number of particles competing for the available condensable vapours, so more vapours condense onto lofted particles, growing them to sizes at which they can act as CCN. This increases the cloud drop number and thus the cloud optical depth. For GEOS-Chem-TOMAS, we ran the same calculation for each month of the year (starting in August 2016), and noticed a seasonal cycle (Extended Data Fig. 9), which is probably because the seasonally dependent position of the Intertropical Convergence Zone shifts the fraction of the tropics that contain pollution emitted in the Northern Hemisphere. Discrepancies between the models and observations exist, and we have considered only the aerosol first indirect effect; therefore these calculations are an estimate only of the magnitude of the radiative effect of tropically upper tropospheric NPF.

## Data availability

The full ATom data set is publicly available[15], as are data specific to this analysis[65].

## Code availability

Code for the model CESM with the base version of CARMA is available online[66] as is code for GEOS-Chem with TOMAS and APM[67]. Code used to analyse ATom data and model output, and recent modifications to CARMA, GEOS-Chem and CAM5 with APM used here, is available on request.

32. Brock, C. A. et al. Aerosol size distributions during the Atmospheric Tomography Mission (ATom): methods, uncertainties, and data products. *Atmos. Meas. Tech*. **12**, 3081–3099 (2019).
33. Williamson, C. et al. Fast time response measurements of particle size distributions in the 3-60 nm size range with the nucleation mode aerosol size spectrometer. *Atmos. Meas. Tech*. **11**, 3491–3509 (2018).
34. Brock, C. A. et al. Sources of particulate matter in the northeastern United States in summer: 2. Evolution of chemical and microphysical properties. *J. Geophys. Res. Atmos*. **113**, D08301 (2008).
35. Brock, C. A. et al. Ultrafine particle size distributions measured in aircraft exhaust plumes. *J. Geophys. Res. Atmos*. **105**, 26555–26567 (2000).
36. Kupc, A., Williamson, C., Wagner, N. L., Richardson, M. & Brock, C. A. Modification, calibration, and performance of the ultra-high sensitivity aerosol spectrometer for particle size distribution and volatility measurements during the Atmospheric Tomography Mission (ATom) airborne campaign. *Atmos. Meas. Tech*. **11**, 369–383 (2018).
37. Cai, Y., Montague, D. C., Mooiweer-Bryan, W. & Deshler, T. Performance characteristics of the ultra high sensitivity aerosol spectrometer for particles between 55 and 800 nm: laboratory and field studies. *J. Aerosol Sci*. **39**, 759–769 (2008).
38. Murphy, D. M. Time offsets and power spectra of the ER-2 data set from the 1987 airborne antarctic ozone experiment. *J. Geophys. Res. Atmos*. **94**, 16737–16748 (1989).
39. Weber, R. J. et al. Spurious aerosol measurements when sampling from aircraft in the vicinity of clouds. *J. Geophys. Res. Atmos*. **103**, 28337–28346 (1998).
40. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* 1st edn (Wiley, 1998).
41. Lovejoy, E. R., Curtius, J. & Froyd, K. D. Atmospheric ion-induced nucleation of sulfuric acid and water. *J. Geophys. Res. Atmos*. **109**, D08204 (2004).
42. Kim, S. et al. PubChem substance and compound databases. *Nucleic Acids Res*. **44**, D1202–D1213 (2016).
43. Westervelt, D. M. et al. Formation and growth of nucleated particles into cloud condensation nuclei: model-measurement comparison. *Atmos. Chem. Phys*. **13**, 7645–7663 (2013).
44. Bowman, K. P. Large-scale isentropic mixing properties of the Antarctic polar vortex from analyzed winds. *J. Geophys. Res. Atmos*. **98**, 23013–23027 (1993).
45. NCAR UCAR Research Data Archive/Computational and Information Systems Lab. *NCEP GFS 0.25-degree global forecast grids historical archive*. https://rda.ucar.edu/datasets/ds084.1/ (2015, updated daily; accessed 28 May 2018).
46. NASA Langley Cloud and Radiation Research. *Satellite imagery and cloud products page* www.satcorps.larc.nasa.gov (accessed 28 May 2018).
47. Murphy, D. M. et al. Single-particle mass spectrometry of tropospheric aerosol particles. *J. Geophys. Res. Atmos*. **111**, D23S32 (2006).
48. DeCarlo, P. F. et al. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Anal. Chem*. **78**, 8281–8289 (2006).
49. Canagaratna, M. R. et al. Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer. *Mass Spectrom. Rev*. **26**, 185–222 (2007).
50. Schroder, J. C. et al. Sources and secondary production of organic aerosols in the northeastern United States during WINTER. *J. Geophys. Res. Atmos*. **123**, 7771–7796 (2018).
51. Stith, J. L. *et al*. An overview of aircraft observations from the Pacific Dust Experiment campaign. *J. Geophys. Res. Atmos*. **114**, D05207 (2009).
52. Kodros, J. K., Cucinotta, R., Ridley, D. A., Wiedinmyer, C. & Pierce, J. R. The aerosol radiative effects of uncontrolled combustion of domestic waste. *Atmos. Chem. Phys*. **16**, 6771–6784 (2016).
53. Yu, F. & Luo, G. Simulation of particle size distribution with a global aerosol model: contribution of nucleation to aerosol and CCN number concentrations. *Atmos. Chem. Phys*. **9**, 7691–7710 (2009).
54. Toon, O. B., Turco, R. P., Westphal, D., Malone, R. & Liu, M. S. A multidimensional model for aerosols: description of computational analogs. *J. Atmos. Sci*. **45**, 2123–2143 (1988).
55. Yu, P. et al. Evaluations of tropospheric aerosol properties simulated by the community earth system model with a sectional aerosol microphysics scheme. *J. Adv. Model. Earth Syst*. **7**, 865–914 (2015).
56. Yu, P. et al. Efficient in-cloud removal of aerosols by deep convection. *Geophys. Res. Lett*. **46**, 1061–1069 (2019).
57. Lamarque, J.F. et al. CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model. *Geosci. Model Dev*. **5**, 369–411 (2012).
58. Luo, G. & Yu, F. Simulation of particle formation and number concentration over the Eastern United States with the WRF-Chem + APM model. *Atmos. Chem. Phys*. **11**, 11521–11533 (2011).
59. Granier, C. et al. *POET, a database of surface emissions of ozone precursors*. http://www.aero.jussieu.fr/projet/ACCENT/POET.php (2005; accessed 23 July 2018).

60. Keller, C. A. et al. HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models. *Geosci. Model Dev.* **7**, 1409–1417 (2014).
61. Lamarque, J. F. et al. Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos. Chem. Phys.* **10**, 7017–7039 (2010).
62. Gong, S. L. A parameterization of sea-salt aerosol source function for sub-and super-micron particles. *Global Biogeochem. Cycles* **17**, 1097 (2003).
63. Mahowald, N. M. et al. Change in atmospheric mineral aerosols in response to climate: last glacial period, preindustrial, modern, and doubled carbon dioxide climates. *J. Geophys. Res.* **111**, D10202 (2006).
64. Kodros, J. K. & Pierce, J. R. Important global and regional differences in cloud-albedo aerosol indirect effect estimates between simulations with and without prognostic aerosol microphysics. *J. Geophys. Res.* **122**, 4003–4018 (2017).
65. Williamson, C. J. et al. *ATom: tropical new particle formation measurements and global-model outputs*. https://doi.org/10.3334/ORNLDAAC/1684 (ORNL DAAC, Oak Ridge, TN, USA;2019).
66. *NCAR/UCAR Community Earth System Model CESM1.0 Public Release*. http://www.cesm.ucar.edu/models/cesm1.0/ (accessed 19 February 2019).
67. *GEOS-Chem Online User's Guide*. http://acmg.seas.harvard.edu/geos/doc/man/ (accessed 19 February 2019).
68. *Natural Earth Free Vector and Raster Map Data*. http://www.naturalearthdata.com (accessed 10 December 2015).

**Acknowledgements** We thank J. Kazil, G. Feingold, T. Goren, D. Fahey and K. Aikin for contributing to this analysis, and the ATom leadership team, science team and crew for contributions to the ATom measurements. We acknowledge support from the US NASA Earth System Science Pathfinder Program under awards NNH15AB12I, NNX15AJ23G, NNX15AH33A and 80NSSC19K0124, and from the US National Oceanic and Atmospheric Administration (NOAA) Health of the Atmosphere and Atmospheric Chemistry, Carbon Cycle, and Climate Programs. A.K. was supported by the Austrian Science Fund's Erwin Schrodinger Fellowship J-3613. J.R.P., J.K.K., K.R.B. and A.L.H. were supported by the US Department of Energy's Atmospheric System Research, an Office of Science, Office of Biological and Environmental Research program, under grants DE-SC0019000 and DE-SC0011780; the US National Science Foundation (NSF), Atmospheric Chemistry program, under grant AGS-1559607; and the NOAA, Office of Science, Office of Atmospheric Chemistry, Carbon Cycle, and Climate Program, under cooperative agreement award NA17OAR430001. G.L. and F.Y. acknowledge funding support from NASA under grant NNX17AG35G and from the NSF under grant AGS-1550816. B.W. and M.D. have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation framework program under grant 640458 (A-LIFE) and from the University of Vienna. P.Y. acknowledges the second Tibetan Plateau Scientific Expedition and Research Program (2019QZKK0604).

**Author contributions** C.J.W., A.K., C.A.B., M.D., B.W., K.D.F., D.M.M., P.C.-J., B.A.N., J.L.J. and T.B. collected the data, and C.J.W. wrote the manuscript with contributions from C.A.B., A.K., J.R.P., K.D.F., D.M.M., P.C.-J., J.L.J. and E.A.R. C.J.W., A.K. and C.A.B. analysed the size distributions. M.D. and B.W. analysed cloud properties. K.D.F. and D.M.M. analysed single-particle compositions, and P.C.-J., B.A.N. and J.L.J. analysed bulk particle composition. J.K.K., A.L.H., K.R.B. and J.R.P. ran GEOS-Chem-TOMAS and J.R.P. developed methods for understanding relevant tropical dynamics. G.L. and F.Y. ran GEOS-Chem-APM and CAM5-APM, and P.Y. ran CESM-CARMA. E.A.R. ran the ATom back trajectories. D.A. and J.C.W. developed the relative differencing method.

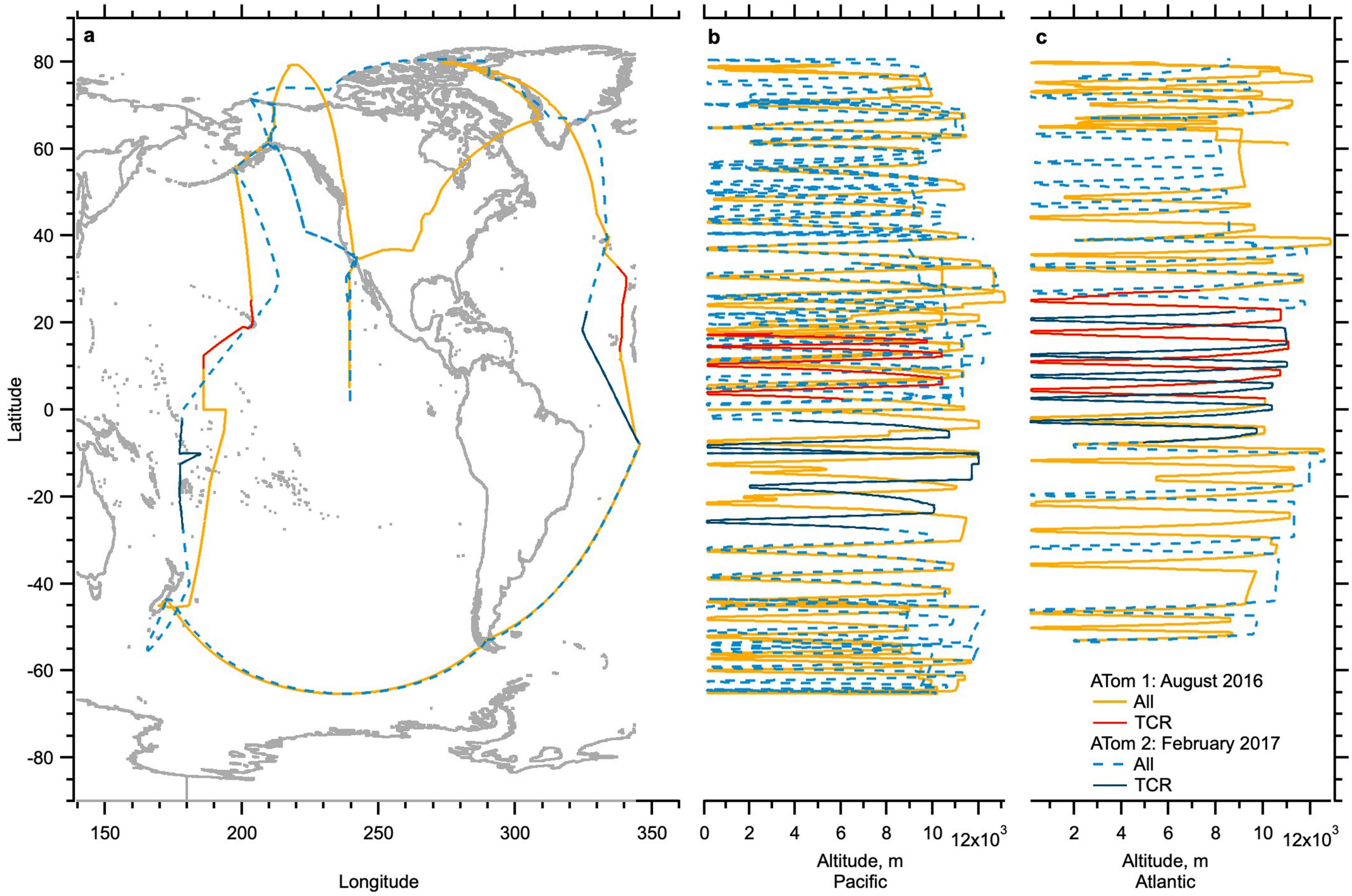
**Competing interests** The authors declare no competing interests.

**Additional information**

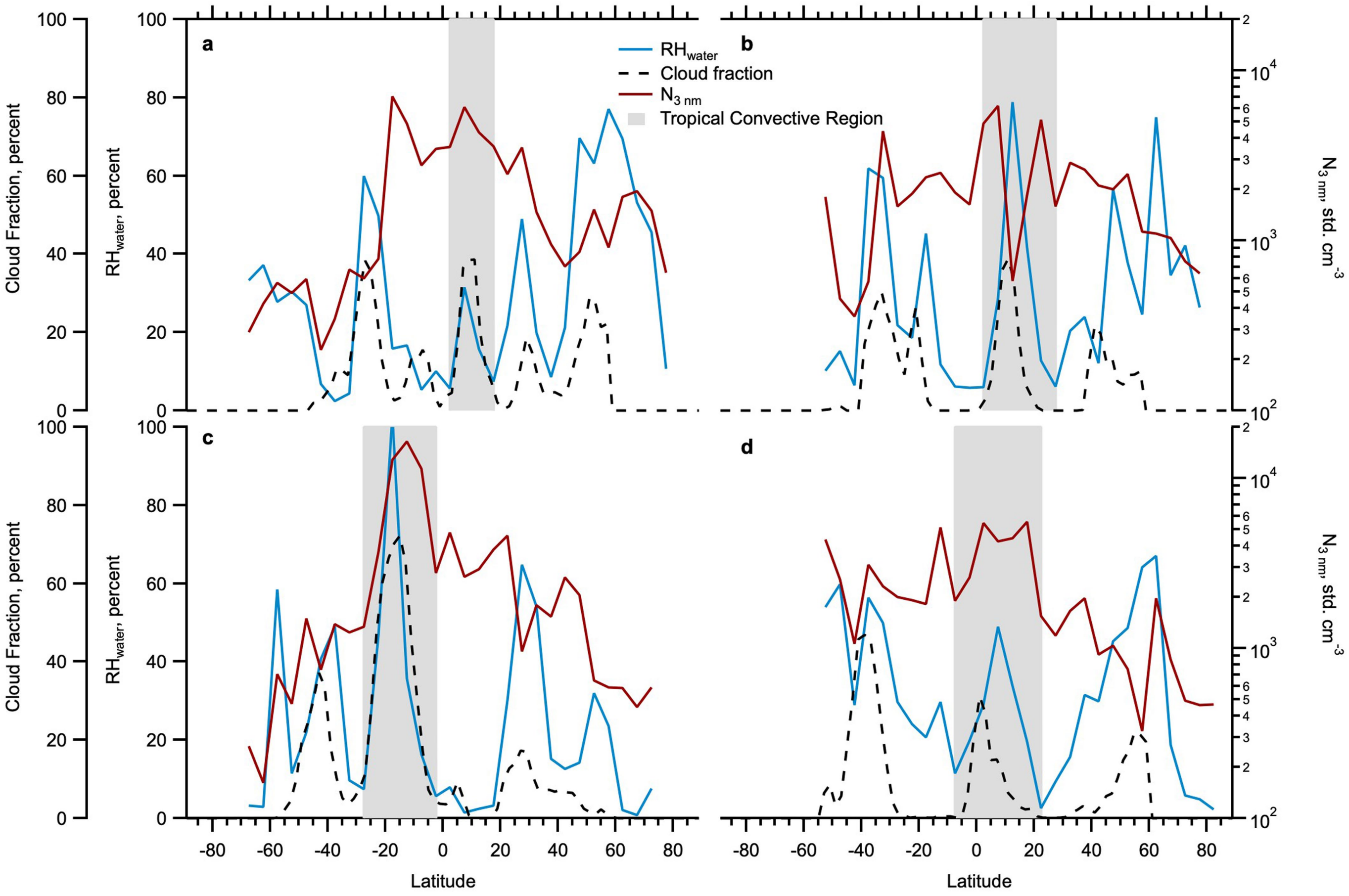
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1638-9.

**Correspondence and requests for materials** should be addressed to C.J.W.

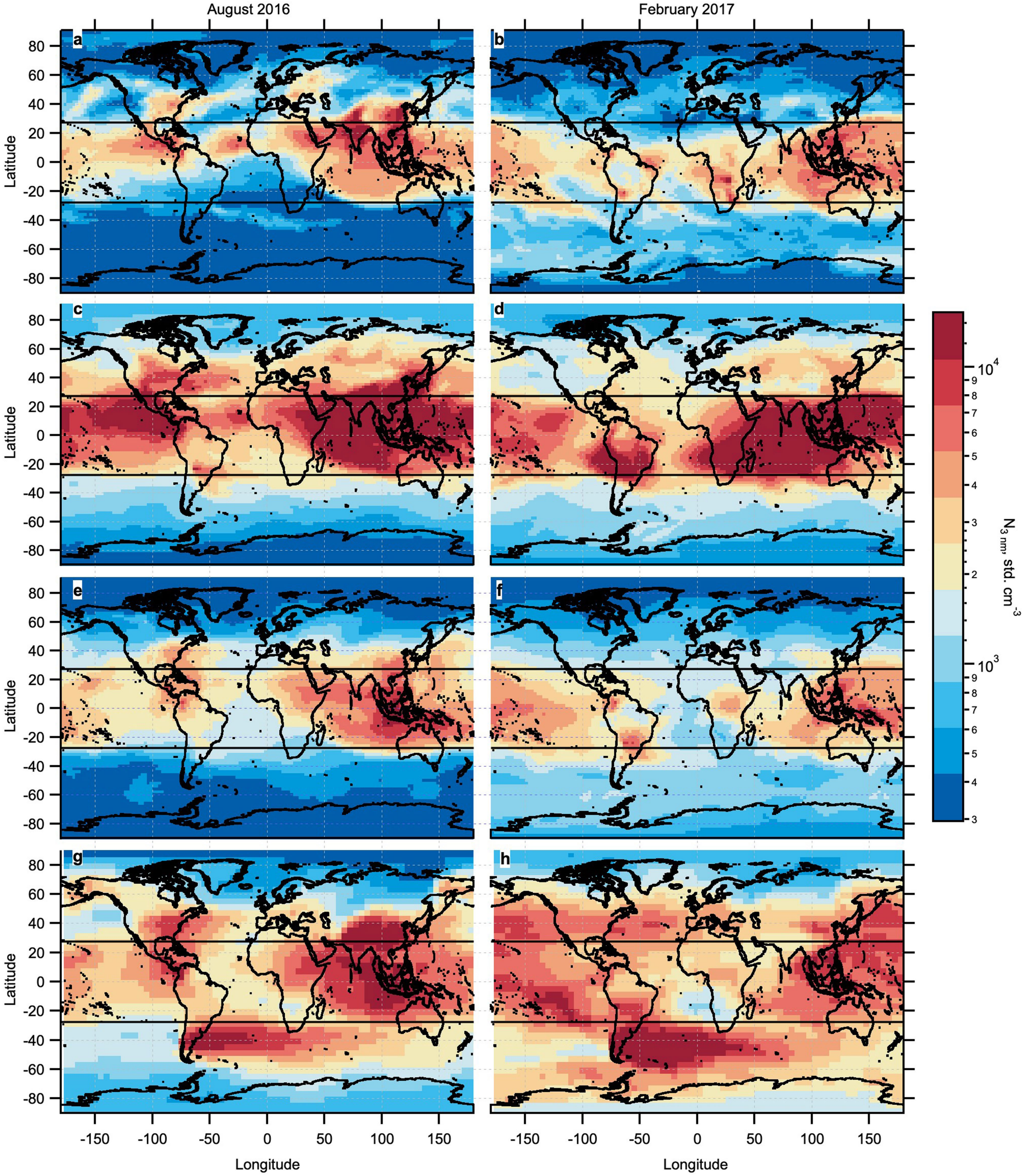
**Peer review information** *Nature* thanks Thorsten Hoffmann and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



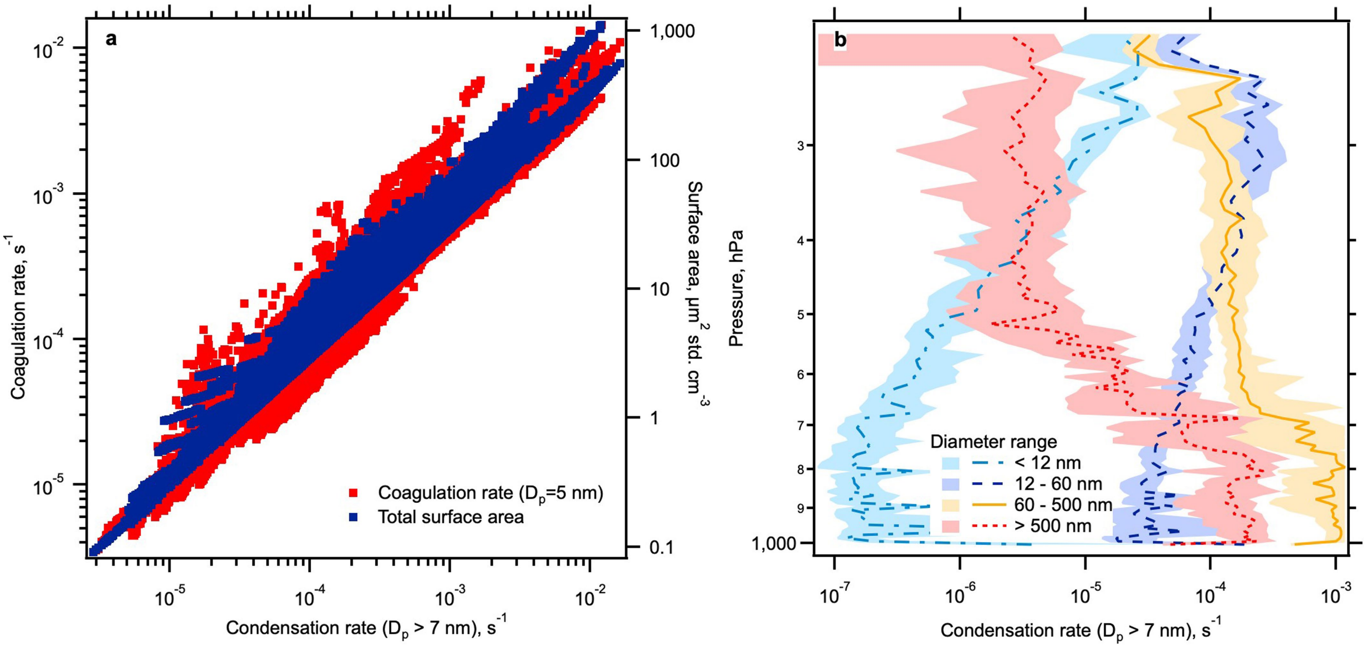
**Extended Data Fig. 1 | Location of ATom measurements. a**, ATom 1 (gold) and ATom 2 (blue) measurements by latitude and longitude. **b**, **c**, Altitude and latitude of measurements over the Pacific Ocean (**b**) and Atlantic Ocean (**c**). TCRs are highlighted in red (for ATom 1) and dark blue (for ATom 2). The world map was made with Natural Earth[68].



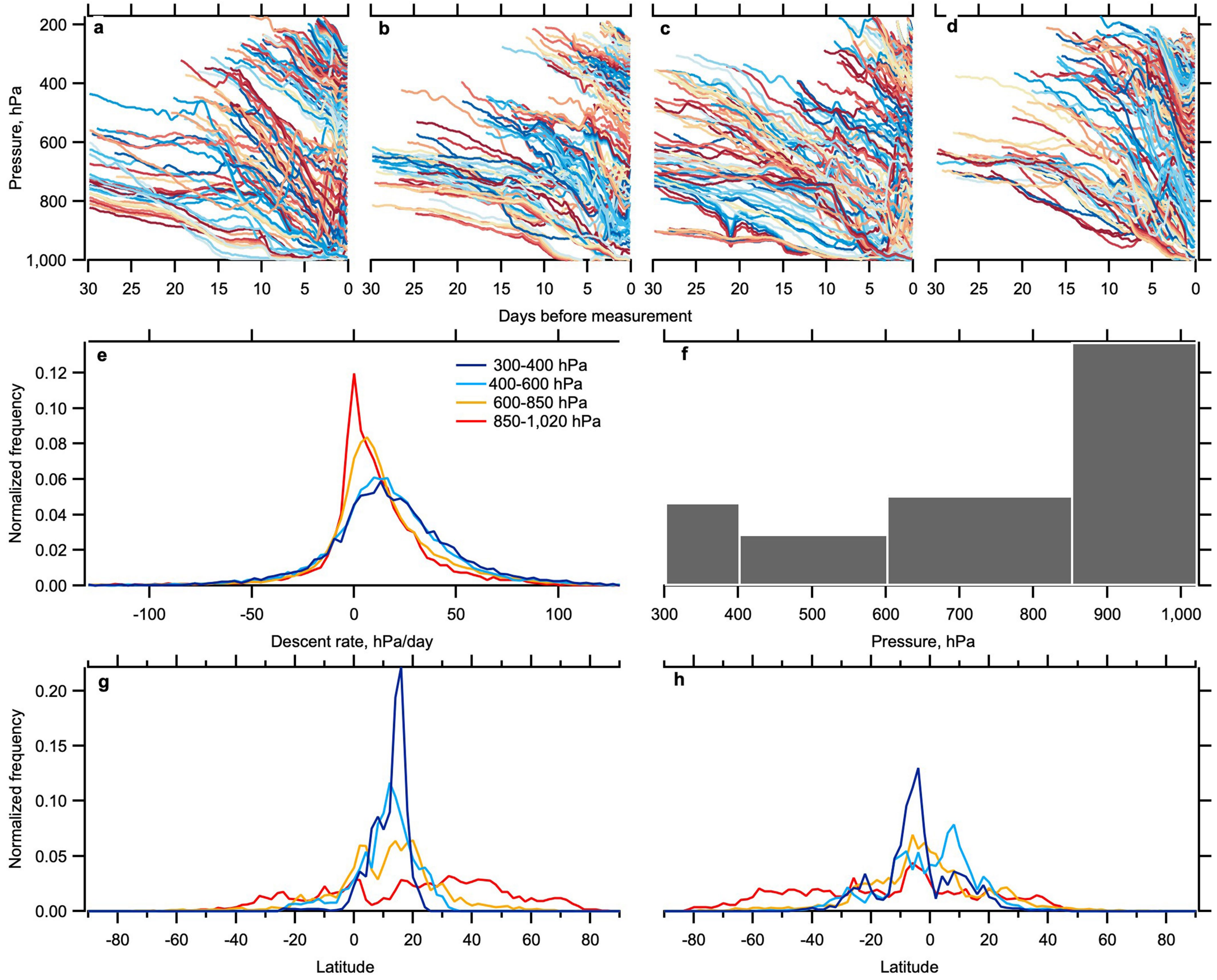
**Extended Data Fig. 2 | Identifying the tropical convective region.** Shown are average measured relative humidity over water (RH$_{water}$; blue); number concentration of particles of 3 nm or more (N$_{3\,nm}$; red); and cloud fraction from reanalysis meteorology (dashed black) at pressures between 200 hPa and 400 hPa. **a**, **b**, ATom 1 Pacific (**a**) and Atlantic (**b**) transects. **c**, **d**, ATom 2 Pacific (**c**) and Atlantic (d) transects. We take the central peak in relative humidity to be the intertropical convergence zone (ITCZ), and define a tropical convective region (TCR) between the minima on either side of this peak (grey shaded region). These minima correspond to latitudes 2.5° N to 17.5° N for ATom 1 Pacific; 2.5° N to 27.5° N for ATom 1 Atlantic; 27.5° S to 2.5° S for ATom 2 Pacific; and 7.5° S to 22.5° S for ATom 2 Atlantic.



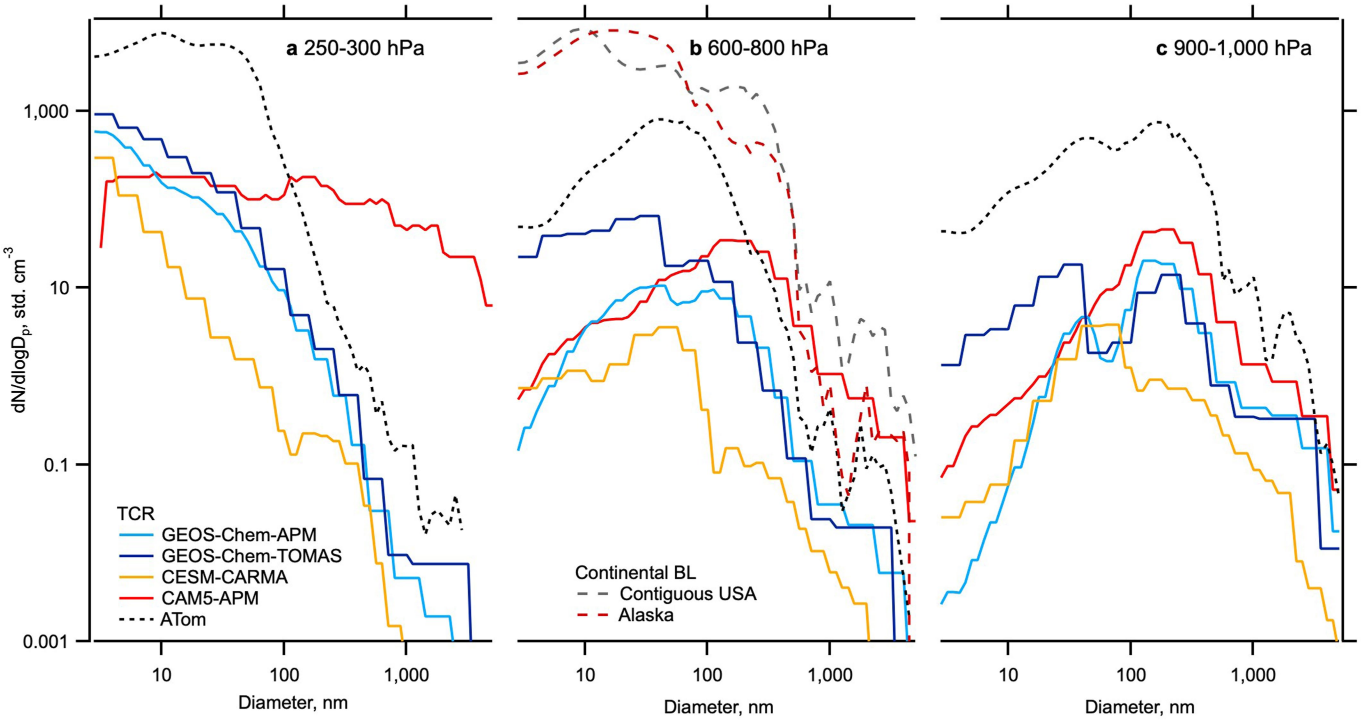
**Extended Data Fig. 3 | Modelled global concentrations of particles larger than 3 nm.** Shown are the monthly mean number concentration of particles bigger than 3 nm in the free troposphere at pressures less than 600 hPa (weighted by grid-box height), modelled for August 2016 (left) and February 2017 (right). **a**, **b**, CESM-CARMA; **c**, **d**, CAM5-APM; **e**, **f**, GEOS-Chem-APM; and **g**, **h**, GEOS-Chem-TOMAS. Horizontal black lines mark the TCR defined by the ATom data. GEOS-Chem-TOMAS (**g**, **h**) shows higher number concentrations of $N_3$ particles outside the TCR than do the other models. This is partly the effect of the 2009 volcanic emissions, which are included in the emissions database for this model. The world maps are made using Natural Earth[68].



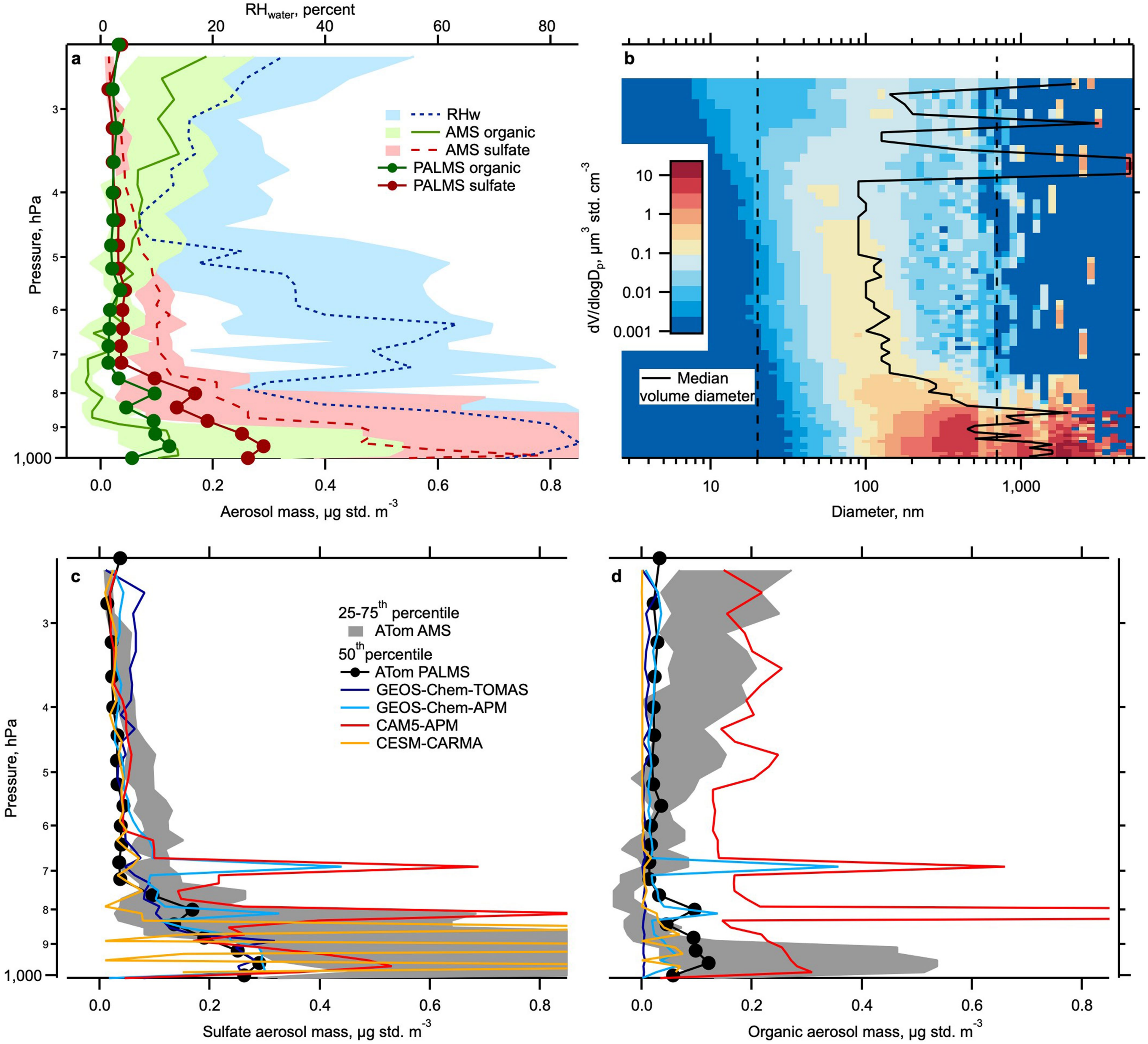
**Extended Data Fig. 4 | Condensation and coagulation rates.** **a**, Relationship between the gas-phase condensation rate onto particles larger than 7 nm, the coagulation rate between 5-nm particles and all other particles, and the total aerosol surface area (2.6–4,800 nm), which show a strongly linear relationship. **b**, The contribution of each particle-size mode to the average condensation rate in the TCR as a function of pressure. The graph shows that the coagulation sink is not always dominated by particles larger than 60 nm, but that, especially at high altitudes, particles of 12–60 nm (the Aitken mode) can dominate the coagulation sink, and must therefore be considered. Lines show the 50th percentile and shaded areas the 25th to 75th percentile range.



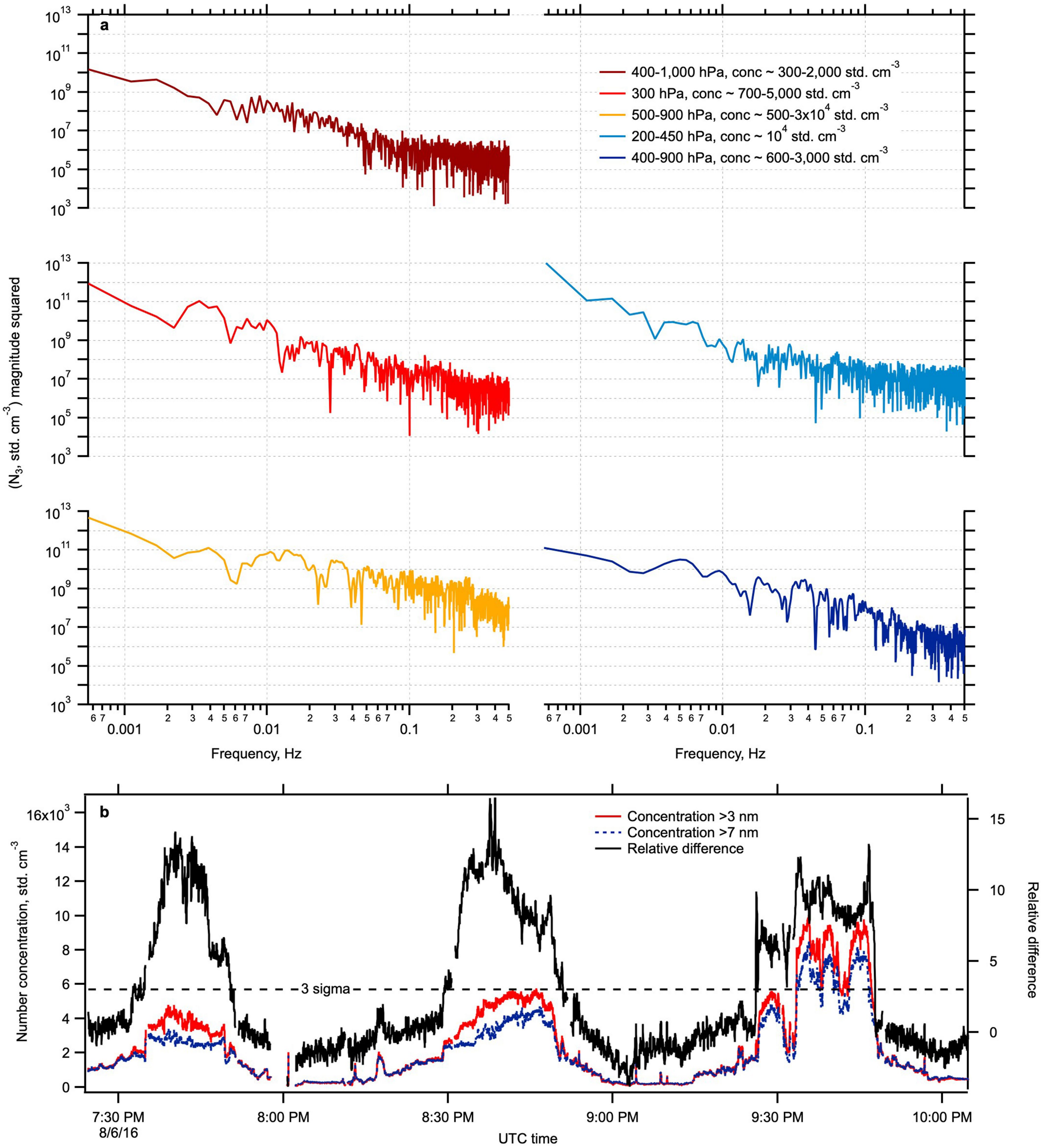
**Extended Data Fig. 5 | TCR back trajectories.** 30-day back trajectories were calculated every minute of flight time within the TCRs. **a**–**d**, Pressures from the time of minimum pressure of the trajectory to the flight track are plotted for all back trajectories from ATom 1 Pacific (**a**) and Atlantic (**b**) and ATom 2 Pacific (**c**) and Atlantic (**d**) observations. Colours distinguish separate trajectories. The general slope of increasing pressure with time indicates a general descending motion of the air. **e**, Histogram showing instantaneous descent rates (one point every 3 h) for all trajectories within the indicated pressure bins. The skew at all altitudes towards positive descent rates is evidence of an overall descending motion of the air. The mean descent rate is higher at higher altitudes, and almost 0 at the lowest altitudes, which is to be expected as this is often within the marine boundary layer where the air cannot descend further. **f**, Average fraction of time that trajectories spent in cloud between the time of minimum pressure and the flight track. In-cloud time is taken to be times when relative humidities are 90% or more (an overestimate). It is binned by the pressure on the flight track (not the pressure of the trajectory itself, as in **e**). For measurements made at pressures of less than 850 hPa, the air spent less than 5% of its time in cloud on average. For air at pressures of more than 850 hPa, this time increased to around 14%. This shows that most of the particles descend with the air instead of being removed by clouds. **g**, **h**, Histograms of the latitudes of trajectories between the minimum pressure and flight track for ATom 1 (**g**) and ATom 2 (**h**), coloured by pressure of the point on the flight track. Except for at the lowest altitudes, air parcels entering the flight track mostly remain within the tropics (with histograms peaking around the equator). Peaks shift towards the summer hemisphere with the season, in the same manner as the TCRs.



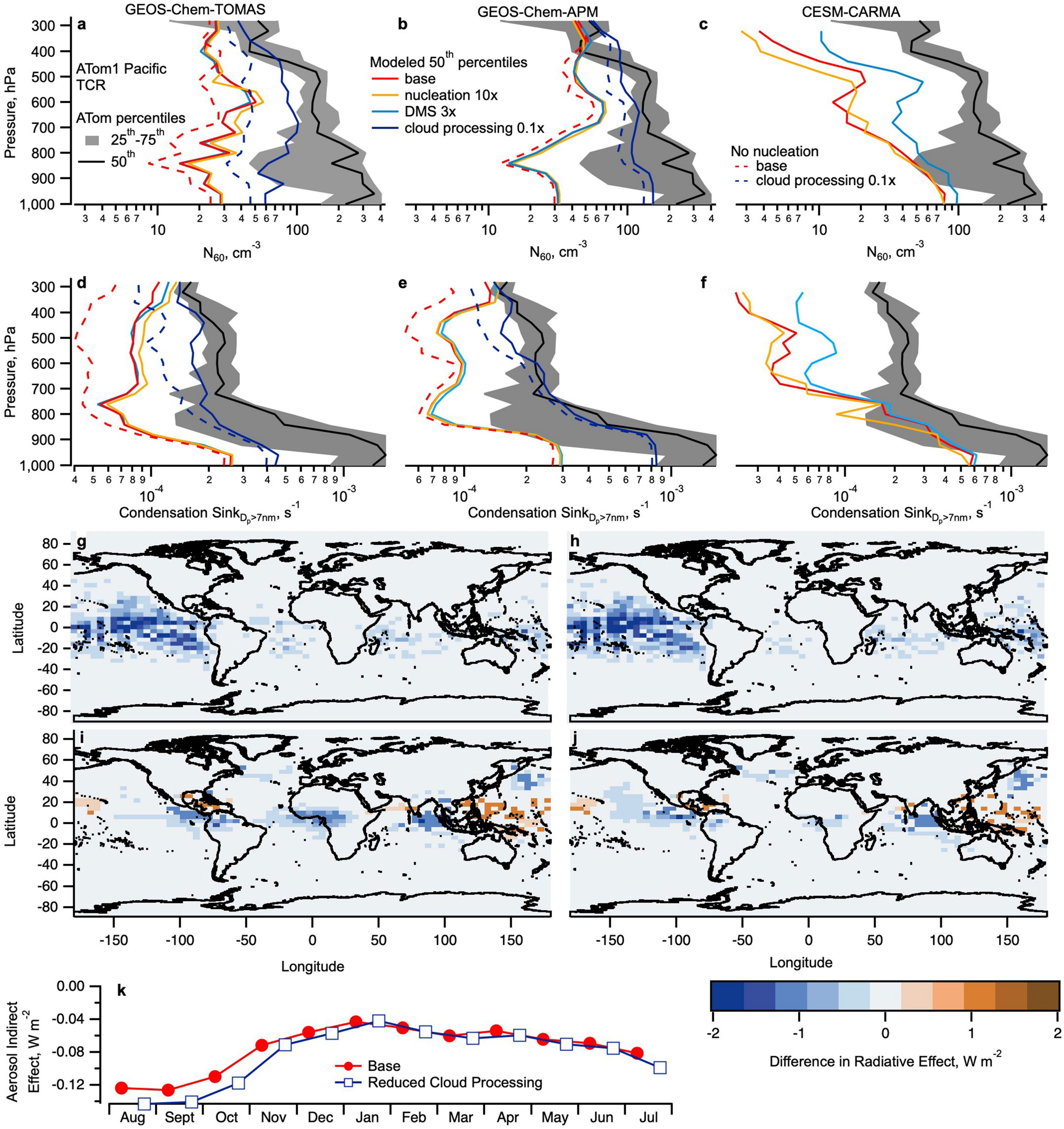
**Extended Data Fig. 6 | Average size distributions. a**–**c**, Average size distributions in TCRs at pressures between 250 hPa and 300 hPa, 600 hPa and 800 hPa, and 800 hPa and 1,000 hPa respectively. Regions of biomass burning or dust plumes have been excluded. Except for particles larger than 100 nm in CAM5-APM, the models show fewer particles than do the ATom observations. All models except for CAM5-APM also show strong evidence of cloud processing, in the form of the dip in the size distribution around 60–100 nm in **b**, **c**. This indication of cloud processing[25] is far less pronounced in the ATom data, suggesting that too many particles are being cloud processed in the models. **b** also shows average size distributions measured between 900 hPa and 1,000 hPa over the contiguous USA and Alaska, as examples of continental size distributions. BL, boundary layer.



**Extended Data Fig. 7 | Chemical composition of particles in the TCRs. a**, Organic and sulfate mass of particles, measured by the AMS (for particles of 50–500 nm) and PALMS (for particles of 150–500 nm), and ambient relative humidity (RH$_w$), with data affected by biomass burning and dust plumes being removed. (The AMS excludes ATom 2 measurements for the Pacific Ocean, where the overall mass was too low to measure sulfate and organic components.) Both methods for measuring organic and sulfate mass have limitations in this regime (the AMS is close to the detection limit, and PALMS cannot measure particles smaller than 150 nm), so perfect agreement is not expected. However, the low sulfate mass at high altitude seems robust, as it is supported by both measurements, and differences between the organic concentrations obtained by PALMS and AMS suggest that organics dominate the composition of smaller particles at high altitudes. **b**, Size-resolved volume in the TCRs. Between 400 hPa and 800 hPa, the median diameter by volume, and the majority of the aerosol volume, is within the measured 50–500-nm range. The composition results should thus be regarded as informative within this pressure range. d$V$/dlog$D_p$, log-normalized volume concentration. **c**, **d**, Modelled sulfate and organic masses over the same regions, compared with measurements.



**Extended Data Fig. 8 | Details from NMASS data. a**, Identifying instrumental noise in particle measurements, using Fourier transforms of the number concentration of particles larger than 3 nm ($N_3$) as measured in the first NMASS channel for 30 min of data at different altitudes and different total concentrations. **b**, Concentrations from the first and second channels of NMASS 1 (red solid and blue dotted lines respectively), and the calculated relative difference (black), for an example time period. Missing data occur when the aircraft flew through a cloud and the data were discarded. A relative difference of three standard deviations (3 sigma) is shown with a dashed horizontal line. Any relative difference larger than this is considered statistically significant.



**Extended Data Fig. 9 | Sensitivity studies and radiative effects.** **a**–**f**, Modelled $N_{60}$ values (**a**–**c**) and condensation sinks (**d**–**f**) in the GEOS-Chem-TOMAS (**a**, **d**), GEOS-Chem-APM (**b**, **e**) and CESM-CARMA (**c**, **f**) models, showing the base model run (red), the nucleation rates increased by a factor of ten between 28° N and 28° S at pressures of less than 600 hPa (orange), oceanic emissions of DMS tripled (light blue), cloud processing on descent reduced by a factor of ten (dark blue), and nucleation turned off between 28° N and 28° S at pressures of less than 600 hPa (dashed). These modelling results are compared with the ATom 1 Pacific observations (black and grey). **g**–**j**, Simulated aerosol indirect effect (difference in radiative effect) of tropical upper tropospheric NPF (**g**, **h**, GEOS-Chem-TOMAS; **i**, **j**, GEOS-Chem-APM), for the base (**g**, **i**) and reduced cloud processing (**h**, **j**) cases, calculated by turning off nucleation at 28° N to 28° S at pressures below 600 hPa. **k**, The seasonal cycle of this global aerosol indirect effect from GEOS-Chem-TOMAS, for the base (red) and reduced cloud processing (blue) cases. Maps made using Natural Earth[68].

**Extended Data Table 1 | Relevant properties of the models used here**

| | Geos-Chem-TOMAS | Geos-Chem-APM | Cam5-APM | CESM-CARMA |
|---|---|---|---|---|
| Meteorology | GEOSFP | MERRA2 | Nudging with MERRA2 | GEOS5 |
| Horizontal Resolution | 4x5 | 2×2.5 | 1.9×2.5 | 1.9×2.5 |
| Vertical Resolution | 47 levels | 47 levels | 32 levels | 56 levels |
| Nucleation Theory | Ternary and binary neutral | Ternary neutral and ternary ion-induced | Ternary neutral and ternary ion-induced | Binary neutral |
| Condensing Gases | $H_2SO_4$, secondary organic vapors* | $H_2SO_4$, low volatility secondary organic gases | $H_2SO_4$, low volatility secondary organic gases | $H_2SO_4$, low volatility secondary organic gases |
| Size Bins (3-60 nm) | 7 | 19 | 19 | 7 |
| Size Bins (> 60 nm) | 8 | 15 | 15 | 13 |
| Mixing State | Internal | Semi-external | Semi-external | Semi-external |

LETTER

https://doi.org/10.1038/s41586-019-1622-4

# Windborne long-distance migration of malaria mosquitoes in the Sahel

Diana L. Huestis[1], Adama Dao[2], Moussa Diallo[2], Zana L. Sanogo[2], Djibril Samake[2], Alpha S. Yaro[2,3], Yossi Ousman[2], Yvonne-Marie Linton[4,5], Asha Krishna[1], Laura Veru[1], Benjamin J. Krajacich[1], Roy Faiman[1], Jenna Florio[1], Jason W. Chapman[6,7,8], Don R. Reynolds[9,10], David Weetman[11], Reed Mitchell[4], Martin J. Donnelly[11], Elijah Talamas[12,13], Lourdes Chamorro[5,12], Ehud Strobach[14,15] & Tovi Lehmann[1]*

**Over the past two decades efforts to control malaria have halved the number of cases globally, yet burdens remain high in much of Africa and the elimination of malaria has not been achieved even in areas where extreme reductions have been sustained, such as South Africa[1,2]. Studies seeking to understand the paradoxical persistence of malaria in areas in which surface water is absent for 3–8 months of the year have suggested that some species of *Anopheles* mosquito use long-distance migration[3]. Here we confirm this hypothesis through aerial sampling of mosquitoes at 40–290 m above ground level and provide—to our knowledge—the first evidence of windborne migration of African malaria vectors, and consequently of the pathogens that they transmit. Ten species, including the primary malaria vector *Anopheles coluzzii*, were identified among 235 anopheline mosquitoes that were captured during 617 nocturnal aerial collections in the Sahel of Mali. Notably, females accounted for more than 80% of all of the mosquitoes that we collected. Of these, 90% had taken a blood meal before their migration, which implies that pathogens are probably transported over long distances by migrating females. The likelihood of capturing *Anopheles* species increased with altitude (the height of the sampling panel above ground level) and during the wet seasons, but variation between years and localities was minimal. Simulated trajectories of mosquito flights indicated that there would be mean nightly displacements of up to 300 km for 9-h flight durations. Annually, the estimated numbers of mosquitoes at altitude that cross a 100-km line perpendicular to the prevailing wind direction included 81,000 *Anopheles gambiae* sensu stricto, 6 million *A. coluzzii* and 44 million *Anopheles squamosus*. These results provide compelling evidence that millions of malaria vectors that have previously fed on blood frequently migrate over hundreds of kilometres, and thus almost certainly spread malaria over these distances. The successful elimination of malaria may therefore depend on whether the sources of migrant vectors can be identified and controlled.**

In Africa, malaria spans the humid equatorial forest to the semi-arid zones in the north and south. In regions in which surface water—which is essential for larval development—is absent during the 3–8-month dry season, mosquito densities and disease transmission drop markedly[3–8]. However, shortly after the first rains, populations of vectors surge[6], and transmission recommences. Recent studies suggest that in the Sahel *A. coluzzii* survives the long dry season by aestivation (a period of dormancy)[3,6,9,10], whereas *A. gambiae* sensu stricto (hereafter, *A. gambiae*) and *Anopheles arabiensis* re-establish populations by migration from distant locations, at which larval sites are perennial[3]. However, direct evidence, such as the capture of aestivating adults in their shelters or the recapture of marked mosquitoes at sites that are hundreds of kilometres from their release sites, remains limited.

The dispersal of mosquitoes (hereafter referred to as migration[11]) has been extensively studied because it directly affects disease transmission, the spread of adaptations (for example, resistance to insecticides) and strategies for controlling mosquitoes (such as insecticide barriers)[12,13]. Although tracking mosquitoes over large scales has seldom been attempted[12,13], the prevailing view is that the dispersal of malaria mosquitoes[12–15] does not exceed 5 km and long-range movements[16–19] represent 'accidental events' that are of minimal epidemiological importance[12]. Nonetheless, the prediction of long-distance migration of anopheline mosquitoes in the Sahel prompted us to question this view. To our knowledge, our study is the first to systematically sample insects that migrate at high altitudes over multiple seasons in Africa. We aimed to determine whether malaria vectors engage in wind-assisted movements and—if so—to assess the epidemiological relevance of this movement by addressing questions regarding the species involved, the frequency and heights of flights, how many mosquitoes migrate and how likely these mosquitoes are to carry *Plasmodium*. We then use simulations to estimate how far mosquitoes may have travelled, and from where.

During 617 nights on which aerial sampling took place, we caught 461,100 insects at heights that ranged between 40 and 290 m above ground level in four villages in the Sahel of Mali (Extended Data Fig. 1). These insects included 2,748 mosquitoes, of which 235 were anopheline mosquitoes (Table 1). These mosquitoes belonged to ten species: *A. coluzzii*, *A. gambiae*, *Anopheles pharoensis*, *Anopheles coustani*, *A. squamosus*, *A. rufipes* and *A. namibiensis*, as well as three distinct but currently undetermined *Anopheles* species (referred to here as *Anopheles* Mali species 1, *Anopheles* Mali species 2 and *Anopheles* species near (sp. nr) *concolor*) (Table 1). *A. coluzzii* and *A. gambiae* are the primary vectors of malaria in Africa, and *A. pharoensis*, *A. coustani*, *A. squamous* and *A. rufipes* are of secondary importance[20]. Mosquitoes were not among the 564 insects that were captured on 508 control nets (Table 1, Methods), which confirmed that these *Anopheles* mosquitoes were intercepted at altitude rather than near the ground during deployment. The maximum number of an anopheline species caught per night was five, which indicates that migration occurred over many nights. Consistent with Poisson distributions, the values of the variance:mean ratio were all near one (Table 1, Supplementary Discussion). Unless specified otherwise, the quantitative results presented here refer to the 5 most-abundant species, represented by more than 20 individuals (Table 1).

---

[1]Laboratory of Malaria and Vector Research, NIAID, NIH, Rockville, MD, USA. [2]Malaria Research and Training Center (MRTC), Faculty of Medicine, Pharmacy and Odonto-stomatology, University of Bamako, Bamako, Mali. [3]Faculte des Sciences et Techniques, Universite des Sciences des Techniques et des Technologies de Bamako (FSTUSTTB), Bamako, Mali. [4]Walter Reed Biosystematics Unit, Smithsonian Institution Museum Support Center, Suitland, MD, USA. [5]Department of Entomology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA. [6]Centre for Ecology and Conservation, University of Exeter, Penryn, UK. [7]College of Plant Protection, Nanjing Agricultural University, Nanjing, China. [8]Environment and Sustainability Institute, University of Exeter, Penryn, UK. [9]Natural Resources Institute, University of Greenwich, Chatham, UK. [10]Rothamsted Research, Harpenden, UK. [11]Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK. [12]Systematic Entomology Laboratory - ARS, USDA, Smithsonian Institution National Museum of Natural History, Washington, DC, USA. [13]Florida Department of Agriculture and Consumer Services, Department of Plant Industry, Gainesville, FL, USA. [14]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA. [15]Global Modeling and Assimilation Office, NASA GSFC, Greenbelt, MD, USA. *e-mail: tlehmann@nih.gov

**Table 1 | Summary of mosquitoes collected in aerial samples in 2013–2015**

| | Standard panels[a] | | | | | | | | | | | Control panels[b] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Taxa | Total captured | Mean panel density | Lower 95% CLM[c] | Upper 95% CLM[c] | Maxium per panel | Per cent nightly presence | Variance: mean ratio | Per cent female | Per cent post-blood feed[d] | Per cent infected[e] | Per cent anthropophily[g] | Total captured | Mean panel density | Maximum per panel |
| *A. squamosus* | 100 | 0.053 | 0.042 | 0.063 | 3 | 11.02 | 1.37 | 76.0 (96) | 93.2 (73) | 0 (73) | 41.1 (17) | 0 | 0 | 0 |
| *A. pharoensis* | 40 | 0.021 | 0.015 | 0.028 | 2 | 6.00 | 1.08 | 82.5 (40) | 100 (33) | 0 (33) | 33.3 (6) | 0 | 0 | 0 |
| *A. coustani* | 30 | 0.016 | 0.01 | 0.022 | 2 | 4.38 | 1.05 | 88.9 (27) | 87.5 (24) | 0 (24) | 14.3 (7) | 0 | 0 | 0 |
| *A. rufipes* | 24 | 0.013 | 0.008 | 0.018 | 2 | 3.24 | 1.16 | 80 (20) | 93.8 (16) | 0 (16) | 0 (4) | 0 | 0 | 0 |
| *A. coluzzii* | 23 | 0.012 | 0.007 | 0.017 | 2 | 3.08 | 1.16 | 95.5 (22) | 90.5 (21) | 0 (21) | 100 (1) | 0 | 0 | 0 |
| *Anopheles*. Mali species 1 | 2 | 0.001 | 0 | 0.003 | 1 | 0.32 | 1 | 100 (2) | 100 (2) | 0 (2) | ND | 0 | 0 | 0 |
| *A. gambiae* | 1 | 0.0005 | 0 | 0.002 | 1 | 0.16 | 1 | 100 (1) | 100 (1) | 0 (1) | ND | 0 | 0 | 0 |
| *Anopheles* sp. nr *concolor*[f] | 1 | 0.0005 | 0 | 0.002 | 1 | 0.16 | 1 | 0 (1) | NA | NA | NA | 0 | 0 | 0 |
| *Anopheles* Mali species 2 | 1 | 0.0005 | 0 | 0.002 | 1 | 0.16 | 1 | 100 (1) | 100 (1) | 0 (1) | ND | 0 | 0 | 0 |
| *A. namibiensis* | 1 | 0.0005 | 0 | 0.002 | 1 | 0.16 | 1 | 100 (1) | 100 (1) | 0 (1) | ND | 0 | 0 | 0 |
| *Anopheles* unidentified | 12 | 0.006 | 0.003 | 0.01 | 1 | 1.78 | 0.99 | 33.3 (6) | 100 (2) | 0 (2) | ND | 0 | 0 | 0 |
| Culicinae | 2,340 | 1.236 | 1.185 | 1.286 | 22 | 58.19 | 4.83 | 86.4 (1,866) | 96.7 (1,629) | ND | ND | 0 | 0 | 0 |
| Culicid unidentified | 173 | 0.091 | 0.078 | 0.105 | 8 | 17.18 | 1.92 | 62.9 (116) | 91.8 (73) | ND | ND | 0 | 0 | 0 |
| Total Culicidae | 2,748 | 1.451 | 1.397 | 1.505 | 23 | 64.18 | 4.92 | 84.5 (1,876) | 96.2 (1,804) | ND | ND | 0 | 0 | 0 |
| Total insects | 461,100 | 243.58 | 242.88 | 244.29 | 2,601 | 100 | 314.75 | ND | ND | NA | NA | 564 | 1.110 | 31 |

Numbers in parentheses refer to the number of insects from which the percentages were calculated (that is, some of the total captured insects could not be assigned a sex or could not be tested). CLM, confidence limit of the mean; NA, not applicable; ND, not determined.

[a] Nightly aerial sampling using sticky nets (panels, usually three per balloon) were launched and retrieved at 17:00 and 07:00, respectively. Nets were raised to set altitudes between 40 and 290 m above ground (Methods). $n = 1{,}894$ panels.

[b] Control panels were raised to 40–120 m above ground level, and immediately retrieved during the launch and retrieval of the standard panels, to estimate the number of insects that were captured during the ascent and descent (Methods). $n = 508$ panels.

[c] These values were estimated using the normal approximation of the Poisson distribution. When a single mosquito per taxon was captured, the low negative values ($< -0.0001$) were rounded to zero.

[d] Only a few blood-fed and half-gravid females were pooled with gravid females to reflect those that were evidently exposed to at least one blood meal. Unfed mosquitoes consisted of the rest.

[e] Infection with human *Plasmodium* species was tested as described in the Methods.

[f] This species was identified on the basis of its genitalia (male).

[g] Identified using PCR (Methods), with additional confirmations by sequencing. The nonhuman hosts included cow, goat and—possibly—unknown rodents.

Females outnumbered males by more than 4:1 (Table 1). Critically, more than 90% of the anopheline females had taken a blood meal (2.9% were blood-fed, 87.5% were fully gravid and 0.7% were semi-gravid; blood feeding is required for the eggs of females to mature) before their high-altitude flights (Table 1), which suggests they were probably exposed to malaria and other pathogens. Although 31% of the blood meals came from humans, no *Plasmodium*-infected mosquitoes were detected among the 22 *A. gambiae* sensu lato (here represented by *A. gambiae* and *A. coluzzii*) that we tested, or the 174 secondary vectors (Table 1). Considering the typical rates of *Plasmodium* infections in primary (1–5%) and secondary (0.1–1%) vectors[5,21–23], our results probably reflect the small sample size: the likelihood for zero infected mosquitoes was over 30% and over 18% (assuming the highest rates in each range) in the primary and secondary vectors, respectively (Supplementary Discussion). Therefore, unless infection reduces migratory capacity or migrants are resistant to parasites (and there is currently no evidence for either), *Plasmodium* and other pathogens are almost certainly transported by windborne mosquitoes that may infect people post-migration.

The mosquitoes were intercepted between 40 and 290 m above ground level (Fig. 1a). The mean panel density of mosquitoes per altitude and the corresponding aerial density (the panel density divided by the volume of air sampled) increased with altitude and there was a significant effect across species on mean panel density ($P < 0.037$, $F_{1,24} = 4.9$) (Extended Data Fig. 2b), which suggests that the migration of anopheline mosquitoes also occurs more than 290 m above ground level. The similar distribution of species across years and locations (Extended Data Fig. 2c; non-significant effects of year and location shown in Extended Data Table 1), combined with the marked seasonality of the migration or high-altitude flight activity (the aerial captures of mosquitoes occurred between July and November, and peaked between August and October) (Fig. 1b, Extended Data Table 1), attest to the regularity of the migration of windborne *Anopheles* mosquitoes.

Using mean aerial densities and wind speeds at altitude (4.8 m s$^{-1}$) (Fig. 1c) and conservatively assuming that mosquitoes fly in a layer between 50 and 250 m above ground level, we estimated the nightly expected numbers of migrants crossing a 1-km line perpendicular to the wind direction. Nightly estimates ranged between 27 (for *A. gambiae*) and 3,719 (for *A. squamosus*) (Fig. 1d). When interpolated over a 100-km line that links our sampling sites (Extended Data Figs. 1a, 2c), annual migrations are estimated to exceed 80,000 *A. gambiae*, 6 million *A. coluzzii* and 44 million *A. squamosus* mosquitoes in that region alone (Fig. 1d). Thus, the migration of windborne mosquitoes in the Sahel occurs on a massive scale.

For each mosquito capture, we estimated the flight trajectories for 2- and 9-h-long flights using the hybrid single-particle Lagrangian integrated trajectory (HYSPLIT) model[24] with the most accurate assimilated meteorological data available (ERA5), and assuming that mosquitoes ascend by their own flight but are passively carried by the wind at altitude (Methods). The mean nightly displacements were 30 and 120 km (maxima of 70 and 295 km) for 2-h and 9-h flights, respectively (Fig. 2, Table 2). Notably, the maximal 9-h nightly flight displacements ranged between 257 and 295 km for all anopheline species with sample size of more than 20 insects (Table 2). These backwards trajectories exhibited a southwestern origin (Rayleigh test; a mean bearing of 212°, $r = 0.54$, $P < 0.0001$) (Table 2), which corresponds to the prevailing winds during peak migration (August to September) (Fig. 2). The trajectories of most species originated from a broad arc (over 90°) (Fig. 2), which suggests migrants emanated from multiple sites across a large region. Migration from this direction is consistent with the presence of high-density populations due to perennial larval sites and earlier population growth following the monsoon rains. The backwards trajectories with a strong northerly component, which were observed during the sparsely sampled period of October to December (Fig. 2), might indicate southward 'return flights' on the Harmattan winds that prevail during this season.

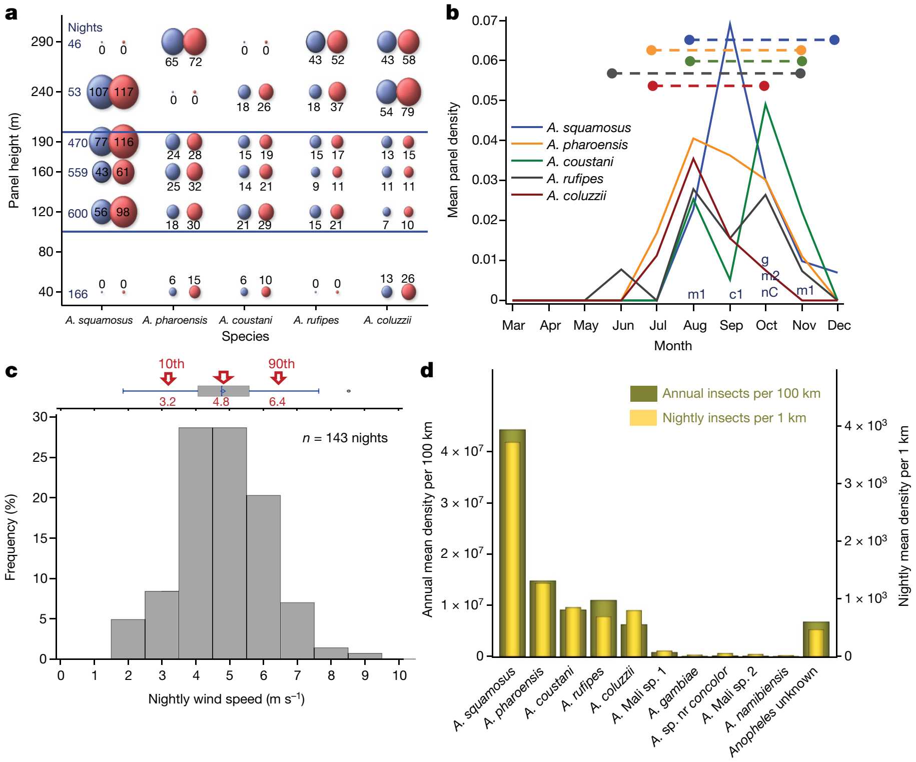


**Fig. 1 | Flight altitude, seasonality, wind speed and abundance of migratory anopheline species. a**, The relationship of altitude (height of the collecting panel) to panel density (blue) and aerial density (orange, mosquitoes per $10^6$ $m^3$ of air), for the five most-common anopheline species (Table 1). Bubble size is proportional to mosquito density (the value shown in the bubble × $10^3$); when the value is zero, only a dot is shown. The number of sampling nights for each of the collecting-panel heights is shown on the left. **b**, Monthly panel density ($n = 1{,}894$ panels) for the five most-common species (Table 1), overlaid by the length of migration period (dashed lines). Values for *A. squamosus* were divided by three to preserve the scale. The sampling month for species that were collected only once or twice is shown by letters: c1, *A. namibensis*; g, *A. gambiae*; m1, *Anopheles* Mali species 1; m2, *Anopheles* Mali species 2; nC, *A.* sp. nr *concolor*. **c**, Distribution of the mean nightly wind speed at flight height on nights on which one or more anopheline mosquitoes were collected. Wind-speed data were taken from ERA5 database after matching the height of the collecting panel to the nearest vertical layer (Methods). The corresponding box and whisker plot (top) shows the median, mean, quartiles and extreme values overlaid by arrows indicating the mean and 10th and 90th percentiles (red). **d**, The number of mosquitoes per species that cross, at altitude (50–250 m above ground level), imaginary lines perpendicular to the prevailing wind direction. Migrants per night per 1 km (right *y* axis) are superimposed on the annual number of migrants per 100-km line (left *y* axis).

Contrary to the conventional view that dispersal of African anopheline mosquitoes occurs over distances[12,14,15,25] of less than 5 km, our results provide compelling evidence that primary and secondary malaria vectors regularly engage in high-altitude flights (or windborne) migrations that span tens to hundreds of kilometres per night. Because this migration includes large numbers of females that had taken at least one blood meal, it probably involves human *Plasmodium* (among other pathogens). Separate outbreaks of malaria in Egypt and Israel have previously been attributed[16] to *A. pharoensis* travelling more than 280 km. Assuming a conservative[22,26] 1% infection rate in migrating females of *A. coluzzii, A. gambiae*, *A. coustani* and *A. pharoensis* and 0.1% in the remaining anopheline mosquitoes (excluding the unknown *Anopheles* Mali species 1 and *Anopheles* Mali species 2) (Supplementary Discussion), more than 286,000 infected migrant mosquitoes are expected to annually cross, at altitude, a 100-km line perpendicular to the prevailing wind direction. Accordingly, *A. pharoensis*, *A. coustani* and *A. coluzzii* would contribute 41%, 25% and 17%, respectively, to malaria transmission by infected windborne mosquitoes. Although these estimates are coarse, this suggests that migratory secondary vectors could be a major source of infections, and that they should be included in studies of transmission and in control programs.

*A. coluzzii* was more common than *A. gambiae* among the migrants, which was contrary to initial predictions[3] that were based on data suggesting that *A. coluzzii* aestivates locally and therefore may not require migration to recolonize the Sahel. Indeed, migration occurs from the end of July to October—well after the surge of *A. coluzzii* populations in the Sahel following the first rain (May to June)[3,6]. Northward and southward oscillations of the intertropical convergence zone during the wet season continually reconfigure the better resource patches for mosquitoes, as the intensity of the rains shifts in location. Additionally, wet-season droughts endanger local mosquito populations every decade or two[27]. Thus, selection pressures to track freshwater resources by riding the winds that bring rain[28] may explain why residents of the Sahel, such as *Oedaleus senegalensis* grasshoppers and *A. coluzzii*, have a mixed strategy of migration[29] and local dormancy. *A. gambiae*, which presumably recolonizes the Sahel every wet season, is relatively rare in villages in the Sahel[3]—therefore, only one specimen of this species was captured by our nets. It may migrate on fewer nights and constitute a smaller fraction of windborne migrants (Supplementary Discussion).

In areas in which malaria is approaching elimination, cases of the disease that occur without travel history are presumed to represent indigenous transmission. We propose that a substantial fraction of such cases, especially those that occur within about 300 km of areas with high rates of malaria transmission, arise from the bites of exogenous, windborne infected mosquitoes. For example, northeastern South Africa has the highest incidence of persistent malaria in the country, with many cases not associated with human travel: these cases are concentrated in an arc that extends over about 150 km from the borders with Zimbabwe and Mozambique, where transmission is high. This area also includes the Kruger National Park, in which roads are scarce and vehicular transport of infected mosquitoes[30] may be hampered. Testing the correlation of infection events such as these with the corresponding prevailing wind

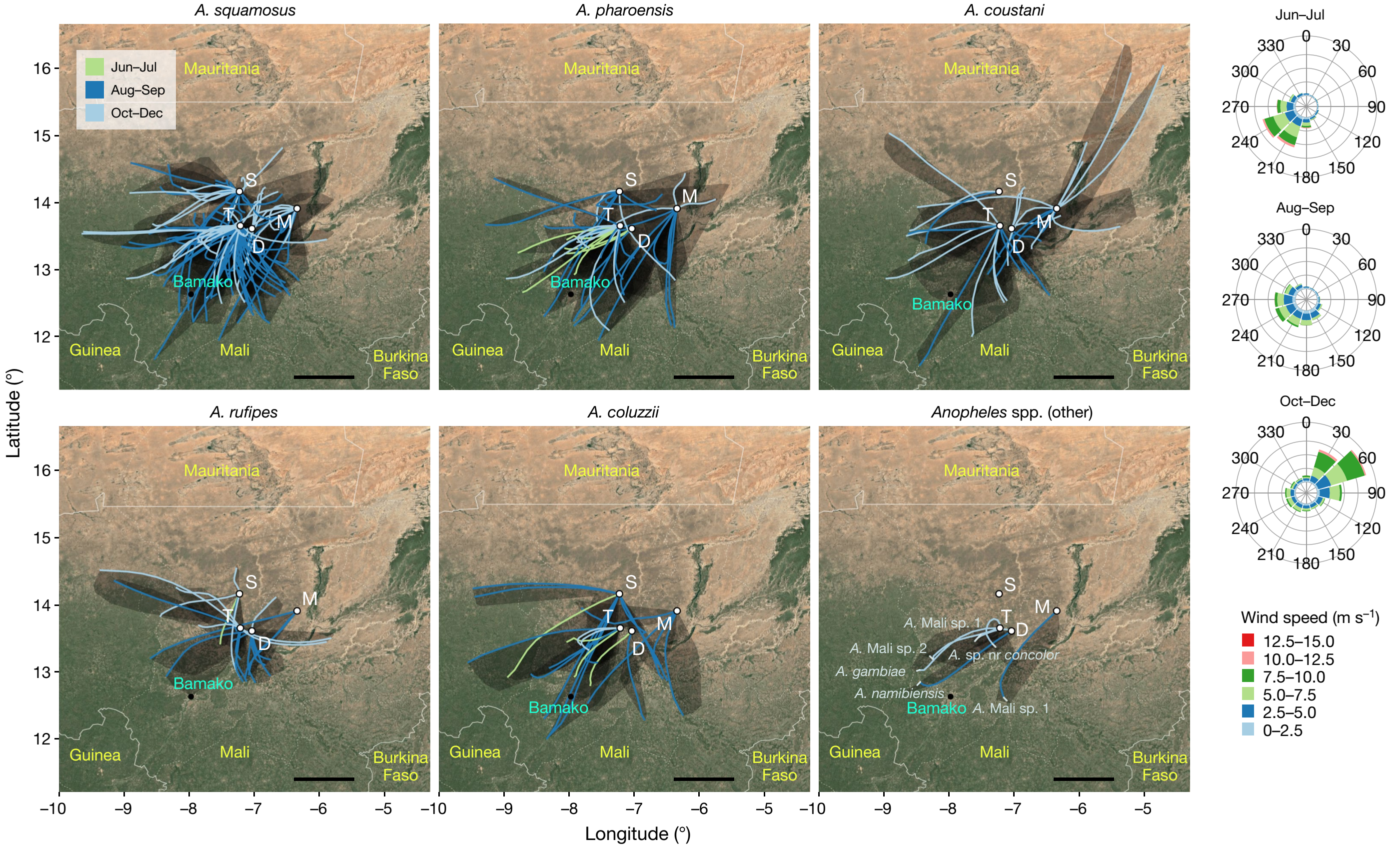


**Fig. 2 | Backwards trajectories of the flights for each *Anopheles* capture event.** Backward 9-h trajectories were estimated by HYSPLIT (Table 2), and are overlaid on a map showing parts of Mali and neighbouring countries. Map data are from Google, Landsat/Copernicus 2019. Each line represents one of four simulated trajectories of one (or more) mosquitoes intercepted at that location and on that night. The area encompassed by the four trajectories is shadowed. Migration season is shown by line colour. The *Anopheles* species is indicated above each panel. D, M, S and T are the balloon launch locations in the villages of Dallowere, Markabougou, Siguima and Thierola, respectively. Scale bars, 100 km. The seasonal-wind rose diagrams, reflecting wind conditions at 180 m above ground level averaged from 2013 to 2015, are shown at the right.

**Table 2 | Summary of displacement distance (in a straight line) and source direction, based on 2-h and 9-h flight trajectories of mosquitoes**

| | Trajectories for 2-h flights | | | | Trajectories for 9-h flights | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Taxa | Trajectories[a] | Displace mean | Displace 95% CLM | Displace min.–max. | Trajectories[a] | Displace mean | Displace 95% CLM | Displace min.–max. | Hourly disp. mean[c] | Actual hourly disp. mean[d] | Mean final bearing[e] | *R* (bearing)[f] | $P_R$ |
| *A. squamosus* | 1,100 | 27.7 | 27–29 | 2–68 | 400 | 109.1 | 103–115 | 4–265 | 13.3 | 12.1 | 213 | 0.516 | 0.0000 |
| *A. pharoensis* | 440 | 31.1 | 30–33 | 1–65 | 160 | 125.3 | 116–134 | 24–260 | 14.7 | 13.9 | 214 | 0.660 | 0.0000 |
| *A. coustani* | 330 | 28.5 | 27–30 | 2–60 | 120 | 125.8 | 114–138 | 16–295 | 14.5 | 14.0 | 199 | 0.270 | 0.0802 |
| *A. rufipes* | 264 | 26.1 | 24–28 | 2–70 | 96 | 109.2 | 97–121 | 24–257 | 12.5 | 12.1 | 199 | 0.454 | 0.0003 |
| *A. coluzzii* | 253 | 38.6 | 37–41 | 3–69 | 92 | 154.1 | 140–168 | 47–270 | 17.3 | 17.1 | 217 | 0.815 | 0.0000 |
| *Anopheles* Mali species 1 | 22 | 20 | 14–26 | 6–52 | 8 | 94.3 | 52–136 | 51–172 | 10.2 | 10.5 | 223 | 0.947 | 0.0000 |
| *A. gambiae* | 11 | 33.5 | ND[b] | ND[b] | 4 | 131.1 | ND[b] | ND[b] | 15.9 | 14.6 | 254 | ND[b] | ND[b] |
| *A.* sp. nr *concolor* | 11 | 17.2 | ND[b] | ND[b] | 4 | 48.2 | ND[b] | ND[b] | 8.4 | 5.4 | 184 | ND[b] | ND[b] |
| *Anopheles* Mali species 2 | 11 | 29.9 | ND[b] | ND[b] | 4 | 104.4 | ND[b] | ND[b] | 13.1 | 11.6 | 234 | ND[b] | ND[b] |
| *A. namibiensis* | 11 | 40.1 | ND[b] | ND[b] | 4 | 149.3 | ND[b] | ND[b] | 16.7 | 16.6 | 241 | ND[b] | ND[b] |
| Overall | 2,453 | 29.4 | 28.8–30.0 | 1–70.4 | 892 | 118.8 | 115–123 | 4–295 | 14.1 | 13.2 | 212 | 0.540 | 0.0000 |

Trajectories were produced using HYSPLIT; further information is provided in Methods and Fig. 2. Disp., displacements; displace mean, mean displacement; displace min.–max., minimum and maximum range of displacement.

[a] The number of unique nightly trajectories assumes all possible nightly interception times, given flight duration and a flight start and end of 18:00 and 06:00, respectively. Thus, for each night with a captured mosquito there were 11 unique 2-h flight trajectories and 4 unique 9-h flight trajectories.

[b] Not determined for species of which only a single specimen was captured.

[c] Hourly displacement between successive 1-h points along the 9-h trajectory.

[d] Effective hourly displacement computed as the quotient of the total 9-h trajectory displacement by 9.

[e] The mean bearing (angle) between the interception point (zero) and the final point of the 9-h trajectory computed from the north.

[f] A measure of angular dispersion that varies from 0 (uniform dispersion from all directions) to 1 (a single angle to which all points align).

direction will help to assess this hypothesis. If confirmed, the incorporation of disease-control efforts in source populations to minimize or block migration is likely to be an essential element in strategies to eliminate malaria.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1622-4.

Received: 15 February 2019; Accepted: 6 September 2019;
Published online 2 October 2019.

1. WHO. *World Malaria Report 2017.* (WHO, 2018).
2. Gething, P. W. et al. Mapping *Plasmodium falciparum* mortality in Africa between 1990 and 2015. *N. Engl. J. Med.* **375**, 2435–2445 (2016).
3. Dao, A. et al. Signatures of aestivation and migration in Sahelian malaria mosquito populations. *Nature* **516**, 387–390 (2014).
4. Fontenille, D. et al. High annual and seasonal variations in malaria transmission by anophelines and vector species composition in Dielmo, a holoendemic area in Senegal. *Am. J. Trop. Med. Hyg.* **56**, 247–253 (1997).
5. Fontenille, D. et al. Four years' entomological study of the transmission of seasonal malaria in Senegal and the bionomics of *Anopheles gambiae* and *A. arabiensis*. *Trans. R. Soc. Trop. Med. Hyg.* **91**, 647–652 (1997).
6. Lehmann, T. et al. Aestivation of the African malaria mosquito, *Anopheles gambiae* in the Sahel. *Am. J. Trop. Med. Hyg.* **83**, 601–606 (2010).
7. Simard, F., Lehmann, T., Lemasson, J. J., Diatta, M. & Fontenille, D. Persistence of *Anopheles arabiensis* during the severe dry season conditions in Senegal: an indirect approach using microsatellite loci. *Insect Mol. Biol.* **9**, 467–479 (2000).
8. Omer, S. M. & Cloudsley-Thompson, J. L. Dry season biology of *Anopheles gambiae* Giles in the Sudan. *Nature* **217**, 879–880 (1968).
9. Mamai, W. et al. Monitoring dry season persistence of *Anopheles gambiae* s.l. populations in a contained semi-field system in southwestern Burkina Faso, West Africa. *J. Med. Entomol.* **53**, 130–138 (2016).
10. Yaro, A. S. et al. Dry season reproductive depression *of Anopheles gambiae* in the Sahel. *J. Insect Physiol.* **58**, 1050–1059 (2012).
11. Chapman, J. W., Reynolds, D. R. & Wilson, K. Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. *Ecol. Lett.* **18**, 287–302 (2015).
12. Service, M. W. Mosquito (Diptera: Culicidae) dispersal—the long and short of it. *J. Med. Entomol.* **34**, 579–588 (1997).
13. Service, M. W. *Mosquito Ecology Field Sampling Methods* (Elsevier Applied Science, 1993).
14. Costantini, C. et al. Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan savanna village. *Med. Vet. Entomol.* **10**, 203–219 (1996).
15. Touré, Y. T. et al. Mark-release-recapture experiments with *Anopheles gambiae* s.l. in Banambani Village, Mali, to determine population size and structure. *Med. Vet. Entomol.* **12**, 74–83 (1998).
16. Garrett-Jones, C. The possibility of active long-distance migrations by *Anopheles pharoensis* Theobald. *Bull. World Health Organ.* **27**, 299–302 (1962).
17. Sellers, R. F. Weather, host and vector—their interplay in the spread of insect-borne animal virus diseases. *J. Hyg. (Lond.)* **85**, 65–102 (1980).
18. Glick, P. A. *The Distribution of Insects, Spiders, and Mites in the Air.* Technical Bulletin No. 673 (US Department of Agriculture, 1939).
19. Reynolds, D. R. et al. Atmospheric transport of mosquitoes in northeast India. *Med. Vet. Entomol.* **10**, 185–186 (1996).
20. Kyalo, D. et al. A geo-coded inventory of anophelines in the Afrotropical Region south of the Sahara: 1898–2016. *Welcome Open Res.* **2**, 57 (2017).
21. Beier, J. C. et al. Characterization of malaria transmission by *Anopheles* (Diptera: Culicidae) in western Kenya in preparation for malaria vaccine trials. *J. Med. Entomol.* **27**, 570–577 (1990).
22. Antonio-Nkondjio, C. et al. Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. *J. Med. Entomol.* **43**, 1215–1221 (2006).
23. Toure, Y. T. et al. Perennial transmission of malaria by the *Anopheles gambiae* complex in a north Sudan Savanna area of Mali. *Med. Vet. Entomol.* **10**, 197–199 (1996).
24. Stein, A. F. et al. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bull. Am. Meteorol. Soc.* **96**, 2059–2077 (2015).
25. Verdonschot, P. F. M. & Besse-Lototskaya, A. A. Flight distance of mosquitoes (Culicidae): a metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. *Limnologica* **45**, 69–79 (2014).
26. Hay, S. I., Rogers, D. J., Toomer, J. F. & Snow, R. W. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, Internet access and review. *Trans. R. Soc. Trop. Med. Hyg.* **94**, 113–127 (2000).
27. Nicholson, S. E. The West African Sahel: a review of recent studies on the rainfall regime and its interannual variability. *ISRN Meteorol.* **2013**, 453521 (2013).
28. Wilson, K. in *Insect Migration: Tracking Resources through Space and Time* (eds Drake, V. A. & Gatehouse, A. G.) 243–264 (Cambridge Univ. Press, 1995).
29. Pedgley, D. E., Reynolds, D. R. & Tatchell, G. M. in *Insect Migration: Tracking Resources through Space and Time* (eds Drake, V. A. & Gatehouse, A. G.) 3–30 (Cambridge Univ. Press, 1995).
30. Frean, J., Brooke, B., Thomas, J. & Blumberg, L. Odyssean malaria outbreaks in Gauteng Province, South Africa, 2007–2013. *S. Afr. Med. J.* **104**, 335–338 (2014).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

**Study area.** Aerial sampling stations were located in four villages in the Sahel in Mali (Extended Data Fig. 1): Thierola (13° 39′ 30.96″ N, 7° 12′ 52.92″ W) from March 2013 to November 2015, Siguima (14° 10′ 3.36″ N, 7° 13′ 40.44″ W) from March 2013 to October 2015; Markabougou (13° 54′ 51.84″ N, 6° 20′ 37.68″ W) from June 2013 to April 2015; and Dallowere (13° 36′ 56.88″ N, 7° 2′ 12.84″ W) from July 2015 to November 2015. This study area has previously been described in detail[3,6,9,11,31–33]. In brief, the region is rural, characterized by scattered villages with traditional mud-brick houses that are surrounded by fields. A single growing season (June to October) enables the farming of millet, sorghum, maize and peanuts, as well as subsistence vegetable gardens. Over 90% of the annual rains fall during this season (about 550 mm). Cattle, sheep and goats graze in the savannah that consists of grasses, shrubs and scattered trees. The rains form small puddles and larger seasonal ponds that usually are totally dry by the end of November. From November until May, rainfall is absent or negligible (total precipitation of less than 50 mm), and by December water is available only in deep wells.

**Aerial sampling and specimen processing.** Aerial sampling stations were placed about 0.5 km from the nearest house of the village in open areas away from large trees. The method of aerial collection of insects was adapted from a study on high-altitude mating flights in ants[34]. Rectangular 3 × 1-m nets (3 $m^2$), cut from a roll of tulle netting (mesh of 8 holes per square centimetre, with hole diameter of 1.2 mm), were sewn to form 4 narrow sleeves, 1 m apart, along the net (Extended Data Fig. 3). A 1-m carbon rod was inserted into each sleeve and glued to the net using Duco Cement Glue (Devcon) (Extended Data Fig. 3). Three nets were spread over each other on a clean large wooden table topped by 3.5 × 1.5-m plywood and coated with a thin film of insect glue (Tanglefoot, Tropical Formula, Contech Enterprises) by rolling a PVC pipe smeared with this glue over them, while applying moderate pressure downward. The pipe was held at each end (from each side of the long table) by two persons and repeatedly rolled (and smeared) until a uniform thin layer of glue coated the net but did not block its holes. After coating, the sticky nets were immediately rolled individually, and kept in two tightly secured plastic bags indoors, to avoid accidental contact with insects before setup.

Before the launch, polyurethane balloons (3 m in diameter; from Mobile Airship & Blimps, or Lighter than Air) were inflated to full capacity with balloon-grade helium (>98.5%) and topped up to ensure full capacity as needed, usually every 1–3 days based on the balloon condition (Extended Data Fig. 3). Typically, balloons were launched over about ten consecutive nights per month. The balloon was kept stationary at about 200 m above ground level by a cord (AmSteelBlue, synthetic rope sling, Southwest Ocean Services) secured to a 1-$m^3$ cement block inserted under the ground. The cord then passed through a horizontal, manually rotating drum made of a garden-hose reel that was used for reeling it in. A larger 3.3-m diameter balloon (Lighter than Air) was used between July and September 2015, and launched to about 300 m above ground level.

A team of five trained technicians operated each aerial-sampling station. During the launch of a balloon, one team member held the cord under the balloon with heavy-duty gloves and manually controlled its ascent and descent, another team member controlled the reel, and the other three team members added or removed the sticky nets to and from their specified positions on the cord. The nets were attached to Velcro panels that had previously been placed on the cord at desirable positions, and spaced to fit each of the matching Velcro pieces on the four carbon rods (Extended Data Fig. 3). A knot was made below the top-most Velcro panel and above the bottom-most Velcro panel, to ensure that the nets would remain stretched (rather than slip on the cord) even in strong winds. Additionally, the team secured the balloons over a 'landing patch' that was padded by tyres covered by a tarpaulin. The balloon was secured to the ground through its main cord by a central hook (at the middle of the landing patch), and by a large tarpaulin that covered it from the top, which was secured to the ground using 14 large stakes. Team members inspected the nets upon launch to verify that they were free of insects. Upon retrieval of the balloon, the team worked in reverse order and immediately rolled each sticky net and placed it into a clean labelled plastic bag, and inserted this into another bag; each of the bags was tightened with a cord until inspection.

Each balloon typically carried three sticky nets. Initially, the nets were suspended at 40, 120 and 160 m above ground level. From August 2013, the typical altitude was set to 90, 120 and 190 m above ground level. When the larger balloon was deployed in the Thierola station (August–September 2015), two additional nets were added at 240 and 290 m above ground level. Balloons were launched approximately 1 h before sunset (about 17:00) and retrieved 1 h after sunrise (about 07:30) the following morning. To control for insects trapped near the ground as the nets were raised and lowered, control nets were raised up to 40 m above ground level and immediately retrieved during the launch and retrieval operations; between September and November 2014, the control nets were raised to 120 m above ground level. The control nets spent 5 min in the air, or up to 10 min when raised to 120 m. Once retrieved, the control nets were processed in a manner identical to that of other nets. After retrieving the panels, inspection for insects was conducted between 09:00 and 11:30 in a dedicated clean area. The panel was stretched between two posts and scanned for mosquitoes, which were counted, removed using forceps and preserved in 80% ethanol before all other insects were similarly processed and placed in other tubes. Depending on their condition, the sticky panels were sometimes reused the subsequent night.

**Species identification.** Glue attached to the insects was washed off with 100% chloroform. The mosquitoes were gently agitated (for less than 30 s) to loosen them from one another. Individual mosquitoes were transferred into consecutive wells filled with 85% ethanol. Using a dissecting scope, the samples were morphologically sorted by mosquito subfamily (Anophelinae or Culicinae), and were tentatively identified to the *Anopheles* species or species-group level. All mosquitoes that were morphologically confirmed as *A. gambiae* sensu lato (and two insects that were identified on the basis of molecular barcode analysis) were identified to the species level on the basis of fragment-size differentiation after amplification of the nuclear ITS2 region and digestion of the product[35]. Validation was carried out in LSTM (laboratory of D.W.), in which each specimen was washed with 500 μl heptane followed by two further washes with ethanol. DNA was then extracted using the Nexttec (Biotechnologie) DNA isolation kit, according to the manufacturer's instructions. Species identification used a standard PCR method, including all primers[36], with products visualized on 2% agarose gel. Samples of *A. gambiae* sensu lato were further identified to species by short interspersed element insertion polymorphism[37]. In cases in which no species-specific bands were detected using the first method, an approximately 800-bp region of the mtDNA cytochrome oxidase I (COI) genes was amplified using the primers C1_J_2183 and TL2_N_3014[38]. PCR products were purified using the QIAquick PCR-Purification kit (Qiagen) and sequenced in both directions using the original PCR primers by MacroGen. Sequences were aligned using CodonCode Aligner (CodonCode) and compared to existing sequences in GenBank to identify species. All other *Anopheles* mosquitoes were identified by the retrospective correlation of DNA barcodes, with morphologically verified reference barcodes compiled by Walter Reed Biosystematics Unit and the Mosquito Barcoding Initiative in the laboratory of Y.-M.L. Head and thorax portions of all samples were separated from abdomens and used for DNA extraction using the Autogen automated DNA extraction protocol. mtDNA COI barcodes were amplified using the universal LCO1490 and HCO2198 barcoding primers[39], and amplified, cleaned and bi-directionally sequenced according to previously detailed conditions[40]. All DNA barcodes generated from this study are available under the project 'MALAN – windborne *Anopheles* migrants in Mali' on the Barcode of Life Database (www.boldsystems.org) and in GenBank under accession numbers MK585944–MK586043. *Plasmodium* infection status was tested for all available *Anopheles* on DNA extracts from the head and thorax portions ($n = 190$) and also from abdomens ($n = 156$) following previously described protocols[41–43]. Owing to the nature of the collections, all body parts were not available for each specimen, which accounts for any discrepancies in the numbers. Blood-meal identification was carried out following a published protocol[44].

**Data analysis.** Although aerial collections started in April 2012, protocol optimization and standardization took most of that year; the data included in the present analysis cover only the period March 2013–November 2015. Nights on which operations were interrupted by storms or strong winds (for example, the balloon was retrieved during darkness) were also excluded.

The total number of mosquitoes per panel represents 'panel density' of each species. Aerial density was estimated on the basis of the panel density of the species, and total air volume that passed through that net that night: that is, aerial density = panel density/volume of air sampled, and volume of air sampled = panel surface area × mean nightly wind speed × sampling duration. The panel surface area was 3 $m^2$. Wind-speed data were obtained from the atmospheric reanalyses of the global climate (ERA5). Hourly data were available at 31-km surface resolution, with multiple vertical levels that included ground, 2, 10, 32, 55, 85, 115, 180, 215, 255 and 300 m above ground level. Overnight records (18:00 through to 06:00) for the nearest grid centre were used to calculate the nightly direction and mean wind speed at each village (Siguima, Markabougou and Thierola). Dallowere, which is located 25 km south of Thierola, was included in the same grid cell as that of Thierola. The mean nightly wind speed at panel height was estimated on the basis of the nearest available altitude layer.

To evaluate clustering in mosquito panel density and the effects of season, panel height, year and locality, mixed linear models with either Poisson or negative-binomial error distributions were implemented by procGLIMMIX[45]. The clustering at the levels of the panel and night of sampling were evaluated as random effects, as was the case for the year of sampling and locality. These models accommodate counts as non-negative integer values. The ratio of the Pearson $\chi^2$ to the degrees of freedom was used to assess the overall 'goodness of fit' of the model, with values of more than two indicating a poor fit. The significance of the scale parameter that estimates $k$ of the negative-binomial distribution was used to choose between Poisson and negative-binomial models. Sequential model fitting was used, starting with random factors before adding fixed effects. Lower Bayesian information criterion values, and the significance of the underlying factors, were also used to select the best-fitting model for each species.

The magnitude of migration of windborne migrants was expressed as the expected minimum number of migrants per species crossing an imaginary line of 1 km perpendicular to the wind direction at altitude. This commonly used measure of abundance assumes that the insects fly in a layer that is 1-km wide, and does not require knowledge of the distance or time the insects fly to or from the interception point[46–48]. We used the mean wind speed at altitude (4.8 m s$^{-1}$; Fig. 1c), and assumed that mosquitoes fly in a layer depth of 200 m between 50 and 250 m above ground level, which conservatively reflects the fact that mosquitoes were captured between 40 and 290 m above ground level (Fig. 1a and Extended Data Fig. 2b). Accordingly, this nightly migration intensity was computed as the product of the mean aerial density across the year (conservatively including periods during which no migrants were captured) by the volume of air passing over the reference line during the night. The corresponding annual index was estimated by multiplying the nightly index by the period of migration of windborne migrants, estimated from the difference between the first and last day and month that a species was captured over the three years. Species that were captured once were assumed to migrate during a single month. The annual number of migrants per species that cross a 100-km line was used because of the similar composition of species across our sampling sites, which spanned 100 km (Extended Data Fig. 1a).

Similar to most insects in their size range[47,49,50], the flight speed of mosquitoes[51,52] does not typically exceed 1 m s$^{-1}$. Because winds at panel altitude attain speeds considerably higher than the speed of the mosquito, flight direction and speed are governed by the wind[46,47] and flight trajectories can therefore be simulated on the basis of the prevailing winds during the night of capture at the relevant locations and altitudes, as has previously been done[53–55]. Accordingly, backward trajectories of mosquito flights were simulated using HYSPLIT[25] on the basis of ERA5 meteorological reanalysis data. Data that are available in ERA5 present the highest spatial and temporal resolution available for this region. Comparisons with data of lower spatial and temporal resolution that are available from the MERRA2 reanalysis data[56], and with the Global Data Assimilation System (available at 0.5° spatial resolution), showed good agreement in trajectory direction and overall distance (data not shown). Trajectories of each captured mosquito were simulated starting at its capture location and altitude, and all multiple interception (full) hours during the night of the collection. Because anopheline mosquitoes are nocturnal, we conservatively assumed that flights started at or after 18:00 and ended by 06:00 the following morning, and computed trajectories for every hour that allowed for a total of a 2-h or 9-h flight. For example, to complete a 9-h flight by 06:00, a mosquito could have started at 18:00, 19:00, 20:00 or 21:00. The total flight duration of tethered female *A. gambiae* sensu lato and *Anopheles atroparvus* reached or exceeded 10 h, with an average speed[51] of 1 km h$^{-1}$ that is consistent with other studies[52,57,58]. Similarly, *Anopheles vagus* and *Anopheles hyrcanus*, caught 150 m above ground level after midnight over India, would have been migrating for more than 6 h, assuming they took off around dusk[20]. Thus, we conservatively assumed that anopheline mosquitoes at high altitude fly between 2 and 9 h per night, although longer durations are possible. Each trajectory consisted of the global positions of the mosquitoes at hourly intervals from the interception time. In addition to plotting trajectories[59–66], the linear distance from the interception site and the simulated position of the mosquito, and the azimuth (angle between the interception site and the simulated position of the mosquito from the north, projected on a plane) were computed for all trajectories. To evaluate the distance range and dominant directions of flight, the mean and 95% confidence interval of the distance and azimuth (as a circular statistic) were computed for the 2-h and 9-h flight trajectories. The dispersion of individual angles (azimuths) around the mean was measured by the mean circular resultant length $r$, which can vary from 0 to 1, with higher values indicating tighter clustering around the mean. Rayleigh's test was used to test whether there was a mean direction; no mean direction occurs when the angles form a uniform distribution over a circle[67].

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Data on anopheline capture, identification, sex and gonotrophic status are available from www.boldsystems.org (project code MALAN) and in GenBank (MK585944–MK586043, inclusive).

## Code availability

SAS code used for statistical analyses (and data manipulations) and 9-h backward trajectories data for each mosquito-capture event (based on HYSPLIT) are available from the corresponding author upon request. The code for plotting trajectories is available at https://github.com/benkraj/anopheles-migration.

31. Lehmann, T. et al. Tracing the origin of the early wet-season *Anopheles coluzzii* in the Sahel. *Evol. Appl*. **10**, 704–717 (2017).
32. Lehmann, T. et al. Seasonal variation in spatial distributions of *Anopheles gambiae* in a Sahelian village: evidence for aestivation. *J. Med. Entomol*. **51**, 27–38 (2014).
33. Huestis, D. L. et al. Seasonal variation in metabolic rate, flight activity and body size of *Anopheles gambiae* in the Sahel. *J. Exp. Biol*. **215**, 2013–2021 (2012).
34. Fritz, G. N., Fritz, A. H. & Vander Meer, R. K. Sampling high-altitude and stratified mating flights of red imported fire ant. *J. Med. Entomol*. **48**, 508–512 (2011).
35. Fanello, C., Santolamazza, F. & della Torre, A. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med. Vet. Entomol*. **16**, 461–464 (2002).
36. Scott, J. A., Brogdon, W. G. & Collins, F. H. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg*. **49**, 520–529 (1993).
37. Santolamazza, F. et al. Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malar. J*. **7**, 163 (2008).
38. Simon, C. et al. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am*. **87**, 651–701 (1994).
39. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol*. **3**, 294–299 (1994).
40. Linton, Y.-M. et al. Mosquitoes of eastern Amazonian Ecuador: biodiversity, bionomics and barcodes. *Mem. Inst. Oswaldo Cruz* **108**, 100–109 (2013).
41. Bass, C. et al. PCR-based detection of *Plasmodium* in *Anopheles* mosquitoes: a comparison of a new high-throughput assay with existing methods. *Malar. J*. **7**, 177 (2008).
42. Demas, A. et al. Applied genomics: data mining reveals species-specific malaria diagnostic targets more sensitive than 18S rRNA. *J. Clin. Microbiol*. **49**, 2411–2418 (2011).
43. Steenkeste, N. et al. Towards high-throughput molecular detection of *Plasmodium*: new approaches and molecular markers. *Malar. J*. **8**, 86 (2009).
44. Kent, R. J. & Norris, D. E. Identification of mammalian blood meals in mosquitoes by a multiplexed polymerase chain reaction targeting cytochrome *b*. *Am. J. Trop. Med. Hyg*. **73**, 336–342 (2005).
45. SAS. SAS for Windows version 9.3 https://www.sas.com/ (2011).
46. Hu, G. et al. Mass seasonal bioflows of high-flying insect migrants. *Science* **354**, 1584–1587 (2016).
47. Drake, V. A. & Reynolds, D. R. *Radar Entomology: Observing Insect Flight and Migration* (CABI International, 2012).
48. Reynolds, D., Chapman, J. & Stewart, A. Windborne migration of Auchenorrhyncha (Hemiptera) over Britain. *Eur. J. Entomol*. **114**, 554–564 (2017).
49. Taylor, L. R. Insect migration, flight periodicity and the boundary layer. *J. Anim. Ecol*. **43**, 225–238 (1974).
50. Chapman, J. W., Drake, V. A. & Reynolds, D. R. Recent insights from radar studies of insect flight. *Annu. Rev. Entomol*. **56**, 337–356 (2011).
51. Kaufmann, C. & Briegel, H. Flight performance of the malaria vectors *Anopheles gambiae* and *Anopheles atroparvus*. *J. Vector Ecol*. **29**, 140–153 (2004).
52. Snow, W. F. Field estimates of the flight speed of some West African mosquitoes. *Ann. Trop. Med. Parasitol*. **74**, 239–242 (1980).
53. Eagles, D., Walker, P. J., Zalucki, M. P. & Durr, P. A. Modelling spatio-temporal patterns of long-distance *Culicoides* dispersal into northern Australia. *Prev. Vet. Med*. **110**, 312–322 (2013).
54. Stefanescu, C., Alarcón, M. & Àvila, A. Migration of the painted lady butterfly, *Vanessa cardui*, to north-eastern Spain is aided by African wind currents. *J. Anim. Ecol*. **76**, 888–898 (2007).
55. Klausner, Z., Fattal, E. & Klement, E. Using synoptic systems' typical wind trajectories for the analysis of potential atmospheric long-distance dispersal of lumpy skin disease virus. *Transbound. Emerg. Dis*. **64**, 398–410 (2017).
56. Gelaro, R. et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim*. **30**, 5419–5454 (2017).
57. Pedgley, D. E. *Windborne Pests and Diseases: Meteorology of Airborne Organisms* (Ellis Horwood, 1982).
58. Gillies, M. T. & Wilkes, T. J. Field experiments with a wind tunnel on the flight speed of some west African mosquitoes (Diptera: Culicidae). *Bull. Entomol. Res*. **71**, 65 (1981).
59. Kahle, D. & Wickham, H. ggmap: spatial visualization with ggplot2. *R J*. **5**, 144–161 (2013).
60. Hijmans, R. J. geosphere: spherical trigonometry. R package version 1.5-10 https://cran.r-project.org/package=geosphere (2017).
61. Slowikowski, K. ggrepel: automatically position non-overlapping text labels with 'ggplot2'. R package version 0.8.1 https://cran.r-project.org/package=ggrepel (2018).
62. Santos Baquero, O. ggsn: north symbols and scale bars for maps created with 'ggplot2' or 'ggmap'. R package version 0.5.2 https://cran.r-project.org/package=ggsn (2019).
63. Arnold, J. B. ggthemes: extra themes, scales and geoms for 'ggplot2'. R package version 4.2.0 https://cran.r-project.org/package=ggthemes (2019).
64. Grolemund, G. & Wickham, H. Dates and times made easy with lubridate. *J. Stat. Softw*. **40**, 1–25 (2011).

65. R Studio Team. RStudio: integrated development environment for R. R package version 3.6.0 https://www.rstudio.com/ (2015).
66. R Core Team. R: a language and environment for statistical computing. https://www.R-project.org/ (2016).
67. Fisher, N. I. *Statistical Analysis of Circular Data* (Cambridge Univ. Press, 1993).
68. Wickham, H. et al. ggplot2: create elegant data visualisations using the grammar of graphics. R package version 3.2.1 https://cran.r-project.org/web/packages/ggplot2/index.html (2019).

**Acknowledgements** We thank the residents of Thierola, Siguima, Markabougou and Dallowere for their consent to work near their homes, and for their wonderful assistance and hospitality; M. Keita, B. Coulibaly and O. Kone for their valuable technical assistance with field and laboratory operations; G. Fritz for consultation on the aerial sampling method using sticky panels; D. Sakai, S. F. Traore, J. Anderson, T. Wellems, M. Sullivan and S. Moretz for logistical support, F. Collins and N. Lobo for support to initiate the aerial-sampling project; J. M. C. Ribeiro and A. Molina-Cruz for reading earlier versions of this manuscript and providing us with helpful suggestions; and A. Crawford and F. (F.) Ngan for conversions of the MERRA2 and ERA5 data files to HYSPLIT format. This study was primarily supported the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. Rothamsted Research received grant-aided support from the UK Biotechnology and Biological Sciences Research Council (BBSRC). Y.-M.L. and R.M. are supported by the US Army. Views expressed here are those of the authors, and in no way reflect the opinions of the US Army or the US Department of Defense. The USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

**Author contributions** The project was conceived by T.L. and D.L.H. Field methods and operations were designed by D.L.H. with input from D.R.R. and J.W.C. Fieldwork, protocol optimization, data acquisition and management, and initial processing of specimens, including tentative species identification, was performed by A.D., A.S.Y., M.D., D.S., Z.L.S. and Y.O. and subsequent processing by A.K., J.F. and L.V. with inputs from E.T. and L.C. Species identification and molecular analysis of specimens were conducted primarily by Y.-M.L., R.M., A.K. and B.J.K. with contributions by D.W., R.F. and M.J.D. Data analysis and HYSPLIT simulations were carried out by T.L. with inputs from all authors, especially R.F., B.J.K., D.R.R., J.W.C., E.S. and Y.-M.L. B.J.K. mapped simulated trajectories. The manuscript was drafted by T.L. and revised by all authors. Throughout the project, all authors have contributed key ideas that have shaped the work and the final paper.

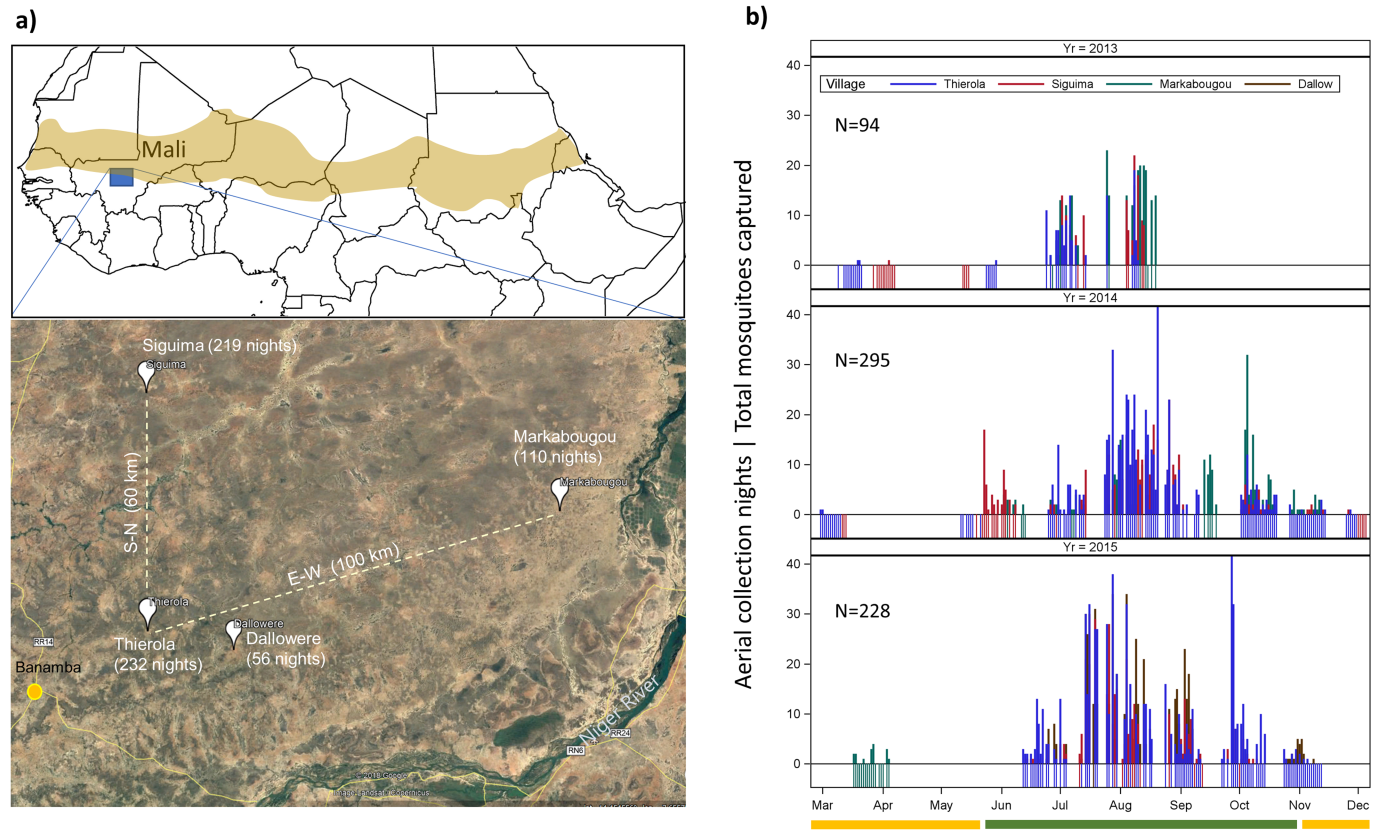
**Competing interests** The authors declare no competing interests.

**Additional information**

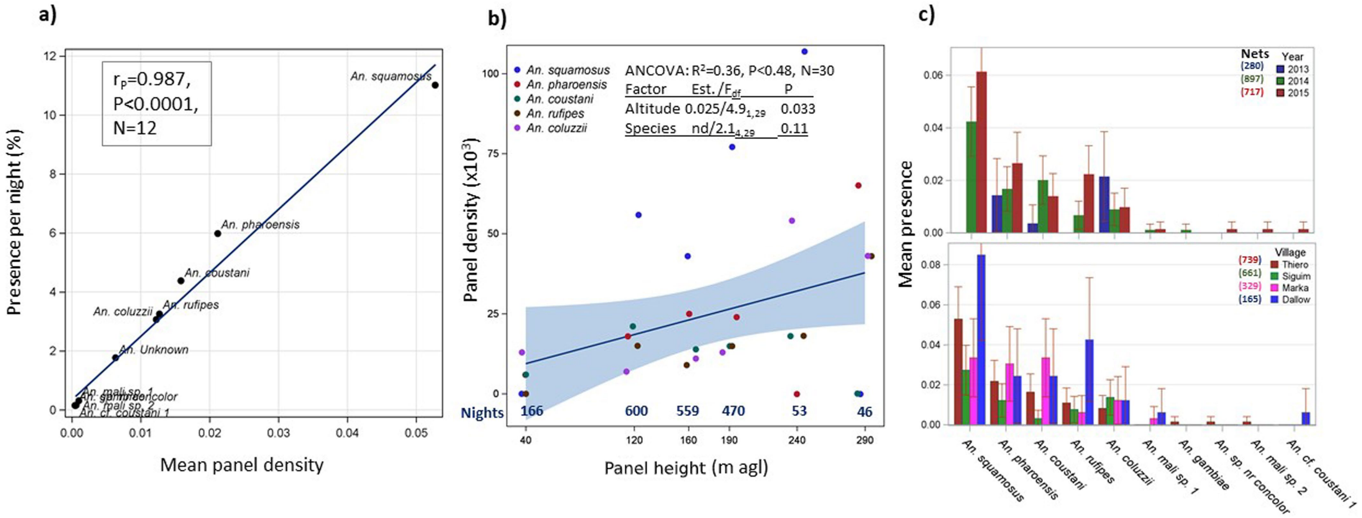
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1622-4.

**Correspondence and requests for materials** should be addressed to T.L.

**Peer review information** *Nature* thanks Nora Besansky, Simon Hay and Daniel Neafsay for their contribution to the peer review of this work.



**Extended Data Fig. 1 | Study area and aerial sampling effort.** **a**, Map of the study area, showing aerial-sampling villages, as well as the number of sampling nights per village. Schematic map of Africa, showing the Sahel region. The base map was generated using the ggplot2 package in R[68], under a GPL-2 license. **b**, Nightly sampling effort by year. The extension of the axes under zero indicates the sampling nights (by village), and the needles denote the total number of mosquitoes per night (regardless of the number of collecting panels per night). Dry and wet seasons are indicated by yellow and green, respectively, in the key under the *x* axis.



**Extended Data Fig. 2 | Regularity of migratory flights, flight altitude and variability among years and localities in the aerial presence of species. a**, Relationship between mosquito presence (fraction of positive nights) and the mean density of mosquitoes on collecting panels, to evaluate whether appearance can be accounted by overall abundance rather than by unique migratory nights. **b**, The relationship between the height of the collecting panel and mean density of mosquitoes per panel ($\times 10^3$; the regression line with shading denotes the 95% confidence interval of the mean), showing the mean density of mosquitoes per panel by species. The inset summarizes the covariance analysis that underlies this regression, which includes the species and height of the collecting panel. The number of nights per collecting-panel height is given in blue along the *x* axis (Fig. 1a). agl, above ground level. **c**, Variation in mosquito presence (fraction of positive nights) by species between years (top) and villages (bottom), with their 95% confidence interval. Sampling effort, expressed as the number of collecting panels per year or village, is shown adjacent to the key.

**Extended Data Fig. 3 | Photograph showing a tethered sticky-panel setup and attachment.** A sticky panel (3 × 1-m net) on a test helium balloon (of a lower volume and capacity), showing the attachment of the net covered with glue to the cord that tethers the balloon to the ground. The four carbon poles and Velcro attachment points are shown. A close-up image of the attachment of the panel to the cord, and an image of preparations to launch a standard 3-m balloon, are also shown.

**Extended Data Table 1 | Variation in the rate of mosquito capture between years, localities and heights above ground**

| **Dependent: Panel Density** | **Parameter** | ***A. squamosus*** | ***A. pharoensis*** | ***A. coustani*** | ***A. rufipes*** | ***A. coluzzii*** |
|---|---|---|---|---|---|---|
| Random vars only: Poisson | Pearson χ2/df (BIC) | 1.13 (793.5) | **1.04 (394.4)** | **0.90 (306.52)** | **1.11 (260.4)** | **1.16 (252.8)** |
| Random vars only: Negative Binomial | Pearson $\chi^2$/df, Scale[a] (BIC) | **0.83, $5.98^{***}$ (756.2)** | 0.97, $3.84^{ne}$ (391.4) | 0.87, $2.09^{ns}$ (306.7) | 0.99, $10.6^{ne}$ (254.5) | 0.98, $15^{ns}$ (246.7) |
| | intercept[mean] (SD) | $-4.06^{ns}$ (1.23) | $-3.9^{**}$ (0.226) | $-4.4^{*}$ (0.63) | $-4.7^{***}$ (0) | $-4.4^{**}$ (0.23) |
| | Year (SD) | $3.24^{ns}$ (4.36) | $0^{ns}$ (0.06) | $0.09^{ns}$ (0.31) | $0.55^{ns}$ (0.56) | $0^{ne}$ |
| | Locality[b] (SD) | $0.075^{ns}$ (0.116) | $0.04^{ns}$ (0.15) | $0.73^{ns}$ (3.19) | $0^{ne}$ | $0^{ne}$ |
| Random vars only: Poisson | Night[c] (SD) | **$4.02^{**}$ (1.42)** | **$1.78^{*}$ (0.99)** | $6.57^{ns}$ (7.3) | **$29.0^{*}$ (16.8)** | **$32.0^{*}$ (17.9)** |
| Random vars only: Neg. Bin. | Night[c] (SD), scale | 3.9** (1.5), $0.74^{ns}$ | $1.6^{ns}$ (1.1), $0.34^{ns}$ | $0.5^{ne}$ (ne), $0^{ne}$ | 30.1* (17.5), $0.7^{ns}$ | 33.5* (18.7), $0.76^{ns}$ |
| Fixed and random: Poisson | Pearson $\chi^2$/df (BIC) | 0.37 (700) | 0.6 (403) | 0.2 (308) | 0.09 (258) | 0.08 (243) |
| | Night | $1.4^{**}$ (0.0) | $0.78^{ns}$ (0.8) | $1.9^{*}$ (1.1) | $14.0^{ns}$(13.3) | $21.9^{ns}$ (15.2) |
| | Period[d] | Aug-$\text{Oct}^{*}$ | Aug-$\text{Oct}^{*}$ | Aug-$\text{Oct}^{ns}$ | Aug-$\text{Oct}^{ns}$ | Aug-$\text{Oct}^{***}$ |
| | Panel height (m) | $0.001^{***}$(0) | $0.003^{***}$(0) | $-0.007^{***}$(0) | $0.001^{***}$(0) | $0.014^{*}$(0.006) |
| **Dependent: Aerial Density** | Pearson $\chi^2$/df (BIC) | 0.42 (938) | 0.41 (503) | 0.2 (378) | 0.1 (304) | 0.09 (283) |
| Fixed and random: Poisson | Night | $2.9^{***}$ (0.8) | $2.6^{*}$ (1.2) | $5.2^{ns}$ (3.9) | $26.8^{*}$(16.0) | $31.5^{*}$ (17.6) |
| | Period[d] | Aug-$\text{Oct}^{ns}$ | Aug-$\text{Oct}^{*}$ | Aug-$\text{Oct}^{ns}$ | Aug-$\text{Oct}^{ns}$ | Aug-$\text{Oct}^{***}$ |
| | Panel height (m) | $-0.003^{***}$(0) | $-0.002^{***}$(0) | $-0.008^{*}$(0.004) | $-0.001^{***}$(0) | $0.01^{*}$(0.005) |

GLIMMIX models of random and fixed variables; total number of panels was 1,894. ***$P = 0.001$, **$P = 0.01$, *$P = 0.05$, NS, not significant ($P > 0.05$); NE, parameter could not be estimated.

[a] The negative-binomial-scale parameter estimates the *k* parameter of this distribution.

[b] When estimating the effect of locality (Extended Data Fig. 1a), only three locations were considered (after pooling Dallowere and Thierola, which are only 20 km apart (Methods)).

[c] The significance of clustering by night (across locations) estimated as the only random effect (using subject statement) after finding insignificant variance components of year and location.

[d] Periods were March to May, June to July, August to October and November to December. The period of the highest panel density is shown, with its statistical significance.

[e] The heights of the panels were 40, 120 (90–120), 160, 190 and 250 (220–290) m above ground level; some heights were combined into ranges, owing to small sample sizes (nights) for particular altitudes.

nature research

Corresponding author(s): Tovi Lehmann

Last updated by author(s): 2019/07/15

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. $F$, $t$, $r$) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's $d$, Pearson's $r$), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was used for primary data collection. ERA5 weather data at altitude were used as part of the data interpretation as detailed in the Methods. |
| Data analysis | Data manipulation and analysis was carried out primarily by SAS 9.4. Flight trajectory simulations were carried out by HYSPLIT 4 using ERA5 weather data. ERA5 weather data were downloaded from the Climate Data Store (https://cds.climate.copernicus.eu/) using Python 3.6.0 scripts, and analyzed with R 3.5.2 (code will be available soon at https://github.com/benkraj/anopheles-migration). Additional R packages, tidyverse, lubridate, ggmap, ggsn, ggthemes, ggrepel, and geosphere were also used in plotting and analysis. All codes used in the analyses relevant to this paper are available as explained in the Data and Code Availability. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability statement is included in out paper: "Data on anopheline capture, identification, sex, and gonotrophic status are available from www.boldsystems.org (Project code: MALAN). Mithochondrial DNA barcode sequences are depsosited in MALAN and in Genbank: MK585944–MK586043). Additionally, 9-hour backward trajectories data for each mosquito capture event based on HYSPLIT are available from TL upon request." These data will be available to the public upon publication of this paper.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | Aerial sampling of insects between 40 and 290 m using nets attached to a stationary helium balloons was performed over 3 years (2013-2015) in four Sahelian villages of Mali. Anopheles mosquitoes collected were identified using morphological keys and molecular barcode analyses. |
| Research sample | Each balloon typically carried three nets. Initially, they were suspended at 40, 120, and 160 m agl, but from August 2013, the typical altitude was set to 90, 120, 190 m agl. When the larger balloon was deployed in the Thierola station (August–September 2015), two additional nets were added at 240 and 290 m agl. Nets tethered to the balloons were launched approximately 1 hour before sunset (~17:00) and retrieved 1 hour after sunrise (~07:30), the following morning. Each panel was stretched between two posts and scanned for mosquitoes, which were counted, removed using forceps, and preserved in 80% ethanol before all other insects were similarly processed and placed in other tubes.<br>To control for insects trapped near the ground as the nets were raised and lowered, control nets were raised up to 40 m agl and immediately retrieved (between September and November 2014 the control nets were raised to 120 m agl) during the launch and retrieval operations. The control nets spent 5 minutes in the air (up to 10 minutes when raised to 120 m). Once retrieved they were processed as other nets. |
| Sampling strategy | We aimed to sample high altitude flying mosquitoes in the Sahel throughout the year for three years to assess regularity of high altitude flight patterns of the species involved. Four sites separated by 25-100 km were compared to assess similarity between locations. Field operations lasted typically two weeks/month (except January and February) to ensure successful sampling for at least 8 nights/month while deferring operation during storms or very windy nights due to major damage to equipment and supplies. However the availability of helium and security concerns, which were outside of our control has affected our operations (e.g., no helium was available for 5 months after August 2013 (despite early and full payment). |
| Data collection | Trained experienced field entomologists supervised, in person the field work in each sampling village. Data, consisting of anopheline and culicine mosquitoes per net per night were recorded on paper and transfered to excel sheets asap by the same team members. |
| Timing and spatial scale | Operations in different villages varied based on availability of helium and security concerns. In Thierola (13.6586, -7.2147) from March 2013 to November 2015, Siguima (14.1676, -7.2279) from March 2013 to October 2015; Markabougou (13.9144, -6.3438) from June 2013 to April 2015; and Dallowere (13.6158, -7.0369) from July 2015 to November 2015. |
| Data exclusions | Aerial sampling started in March 2012. However, protocol optimization took most of this year. Therefore, only data collected after standard operations protocols have been formalized were used. Additionally, sampling nights with strong winds and storms have been excluded because either balloon launch or retrieval were affected. |
| Reproducibility | To ensure reproducibility we only used data obtained after protocol optimization, ie., since March 2013 (ie., we excluded data collected in 2012 see above). To control for insects trapped near the ground as the nets were raised and lowered, control nets were raised up to 40 m agl and immediately retrieved (between September and November 2014 the control nets were raised to 120 m agl) during the launch and retrieval operations. The control nets spent 5 minutes in the air (up to 10 minutes when raised to 120 m). Once retrieved they were processed as other nets. Slight changes in panel altitude were introduced after noticing a trend of higher mosquito catch with altitude as described in Methods. |
| Randomization | NA. All data were used (after exclusion of data prior to 2013 and during storms and strong winds, as noted above) |
| Blinding | Blinding was not intentionally done. However, ultimate mosquito identification using molecular barcode analysis was carried out by collaborators not involved in the field operations. |

Did the study involve field work? ☒ Yes ☐ No

## Field work, collection and transport

| | |
|---|---|
| Field conditions | Field operations were carried out in the Sahel of Mali. The region is rural, characterized by scattered villages with traditional mud-brick houses, surrounded by fields. A single growing season (June–October) allows the farming of millet, sorghum, maize, and peanuts, as well as subsistence vegetable gardens. Over 90% of the annual rains fall during this season (~550mm). Cattle, sheep, and goats graze in the savannah that consists of grasses, shrubs, and scattered trees. The rains form small puddles and larger seasonal ponds that usually are totally dry by the end of November. From November until May, rainfall is absent or negligible (total precipitation < 50mm), and by December through most of May, water is available only in deep wells. Temperatures peak (night high >30C) during the late dry season (March-May) and reach their minima (night near 15-20C) during the early dry season (December-February). Troughout the dry season (December-May) relative humidity is low (25-15%) with few short exceptions, whereas during the wet season (June-November) it is higher (60-90%) based on the monsoon rains. In the |

| | |
|---|---|
| | rest of the year, the temperatures are intermediate and depend on monsoon rains. No electricity or water pipes are available in the villages near our aerial sampling stations. |
| Location | Sampling stations were in Thierola (13.6586, -7.2147), Siguima (14.1676, -7.2279), Markabougou (13.9144, -6.3438), and Dallowere (13.6158, -7.0369). |
| Access and import/export | All the work was coordinated and carried out by the Malaria Research and Training Center of Mali (MRTCH, now the Mali ICEMR). Specimens were exchanges based on existing agreements for scientific cooperation between MRTC, NIH and participating collaborators. Dead and preserved insects were transfered between countries following existing procedures. |
| Disturbance | We believe that aerial sampling using small nets had minimal effect on diversity and biological systems being sampled given the fraction of the space actually sampled. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | Antibodies |
| ☒ | ☐ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology |
| ☐ | ☒ | Animals and other organisms |
| ☒ | ☐ | Human research participants |
| ☒ | ☐ | Clinical data |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | ChIP-seq |
| ☒ | ☐ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

# Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | The study did not involve laboratory animals |
| Wild animals | Only anopheles mosquitoes are included in this paper. Other insects were also collected and be summarized separaetly. |
| Field-collected samples | Almost all insects collected by sticky nets were preserved immediately in 80% ethanol. They were kept in field conditions for several days before they were moved to a laboratory freezer in Bamako, Mali. |
| Ethics oversight | No ethical approval is required for sampling aerial flying mosquitoes. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

https://doi.org/10.1038/s41586-019-1610-8

# Genome editing retraces the evolution of toxin resistance in the monarch butterfly

Marianthi Karageorgi[1,8], Simon C. Groen[1,2,8], Fidan Sumbul[3], Julianne N. Pelaez[1], Kirsten I. Verster[1], Jessica M. Aguilar[1], Amy P. Hastings[4], Susan L. Bernstein[1], Teruyuki Matsunaga[1], Michael Astourian[1], Geno Guerra[5], Felix Rico[3], Susanne Dobler[6], Anurag A. Agrawal[4,7] & Noah K. Whiteman[1]*

**Identifying the genetic mechanisms of adaptation requires the elucidation of links between the evolution of DNA sequence, phenotype, and fitness[1]. Convergent evolution can be used as a guide to identify candidate mutations that underlie adaptive traits[2–4], and new genome editing technology is facilitating functional validation of these mutations in whole organisms[1,5]. We combined these approaches to study a classic case of convergence in insects from six orders, including the monarch butterfly (*Danaus plexippus*), that have independently evolved to colonize plants that produce cardiac glycoside toxins[6–11]. Many of these insects evolved parallel amino acid substitutions in the α-subunit (ATPα) of the sodium pump ($Na^+/K^+$-ATPase)[7–11], the physiological target of cardiac glycosides[12]. Here we describe mutational paths involving three repeatedly changing amino acid sites (111, 119 and 122) in ATPα that are associated with cardiac glycoside specialization[13,14]. We then performed CRISPR–Cas9 base editing on the native *Atpα* gene in *Drosophila melanogaster* flies and retraced the mutational path taken across the monarch lineage[11,15]. We show in vivo, in vitro and in silico that the path conferred resistance and target-site insensitivity to cardiac glycosides[16], culminating in triple mutant 'monarch flies' that were as insensitive to cardiac glycosides as monarch butterflies. 'Monarch flies' retained small amounts of cardiac glycosides through metamorphosis, a trait that has been optimized in monarch butterflies to deter predators[17–19]. The order in which the substitutions evolved was explained by amelioration of antagonistic pleiotropy through epistasis[13,14,20–22]. Our study illuminates how the monarch butterfly evolved resistance to a class of plant toxins, eventually becoming unpalatable, and changing the nature of species interactions within ecological communities[2,6–11,15,17–19].**

Convergently evolved substitutions in ATPα have been hypothesized to contribute to cardiac glycoside resistance in the monarch butterfly and other specialized insects via target-site insensitivity (TSI) in the sodium pump[6–11]. However, it is unclear whether the changes are sufficient for resistance in whole organisms[6–11,15,18,23] or are 'molecular spandrels'—candidate adaptive alleles that do not confer a fitness advantage when tested more rigorously[1,5]. In addition, the evolutionary order of substitutions suggests a constrained adaptive walk[11,13,14,20–22,24], but an in vivo genetic dissection has not been conducted, so it is not possible to draw conclusions about the adaptive role of these substitutions[1–5,15].

We have identified a core set of amino acid substitutions in cardiac glycoside-specialized insects that define potential mutational paths to resistance and TSI. We focused on the first extracellular loop (H1–H2) of ATPα, where most candidate TSI-conferring substitutions occur[7–11] (Fig. 1a). We used maximum likelihood to reconstruct ancestral states for cardiac glycoside specialization (feeding and sequestering) and amino acids within the H1–H2 loop of ATPα across a species phylogeny (Fig. 1b, Methods, Supplementary Text, Supplementary Tables 1–3). Sites 111 and 122 underwent frequent parallel substitutions associated with specialization[7–11] (Fig. 1b, Extended Data Fig. 1). In addition, site 119 experienced repeated substitutions in specialized insects, and co-evolved with site 111. However, substitutions at site 119 were not associated with specialization, because they also occurred in non-specialized insects (Fig. 1b, Extended Data Fig. 1). To determine whether substitutions at sites 111 or 122 appeared in an ordered fashion relative to substitutions at site 119, we compared the mutational order in 21 specialized lineages to a random permutation null model (Fig. 1c), and found that the ordering was unlikely to have occurred by chance (Fig. 1d). A replacement at site 119 always occurred before or with a replacement at site 122, and repeated substitutions at the three sites evolved concurrently with specialization (Fig. 1b).

The mutational paths lead to three predictions for how substitutions at sites 111, 119 and 122 affect fitness. First, the mutational paths provide stepwise fitness advantages at increasing toxin concentrations. Second, the mutational paths contribute to sequestration of cardiac glycosides through passive toxin accumulation. Third, given the ordered appearance of the substitutions, interactions between substitutions (epistasis) increase fitness and mitigate the pleiotropic fitness costs of adaptive substitutions. We focused on the mutational path taken by the monarch lineage, which includes species that do not feed on cardiac glycoside-producing plants and those that sequester the toxins[11] (Fig. 2a).

We used CRISPR–Cas9 genome editing coupled with homology-directed repair (HDR) to generate viable, homozygous *Atpα* knock-in *Drosophila* lines carrying the precise substitutions at sites 111, 119 and 122 of 4 consecutive *Atpα* genotypes in the monarch lineage: LAN, LSN, VSN, and VSH[11,25,26] (substitutions Q111L, A119S, L111V, and N122H, respectively; Fig. 2b, Methods and Supplementary Tables 4–7). We also generated control lines in the same genetic background, and viable, homozygous lines carrying the key single substitutions A119S (QSN) and N122H (QAH), which evolved along the monarch lineage mutational path, but not alone (Fig. 2b, Extended Data Fig. 2). These lines did not carry exogenous DNA sequences or other non-synonymous mutations in the edited region (Supplementary Table 6). We found substantial variation for egg–adult survival among the lines (Extended Data Fig. 3), but *Atpα* mRNA levels and baseline sodium pump activity were not detectably altered (Extended Data Fig. 3, Supplementary Table 8). A congener of *Drosophila melanogaster*, *D. subobscura*, is naturally segregating at positions 111, 119 and 122 for the monarch butterfly genotype (VSH) and was reared from cardiac glycoside-producing plants, indicating that *Drosophila* is a reasonable model for studying cardiac glycoside resistance and TSI[27,28].

We obtained in vivo evidence for adaptation in monarch lineage *Atpα* through larval–adult and adult survival experiments. Knock-in fly lines were reared on yeast medium with increasing concentrations

---

[1]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY, USA. [3]LAI, U1067 Aix-Marseille Université, Inserm, CNRS, Marseille, France. [4]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA. [5]Department of Statistics, University of California, Berkeley, Berkeley, CA, USA. [6]Molecular Evolutionary Biology, Zoological Institute, Biocenter Grindel, Universität Hamburg, Hamburg, Germany. [7]Department of Entomology, Cornell University, Ithaca, NY, USA. [8]These authors contributed equally: Marianthi Karageorgi, Simon C. Groen. *e-mail: whiteman@berkeley.edu

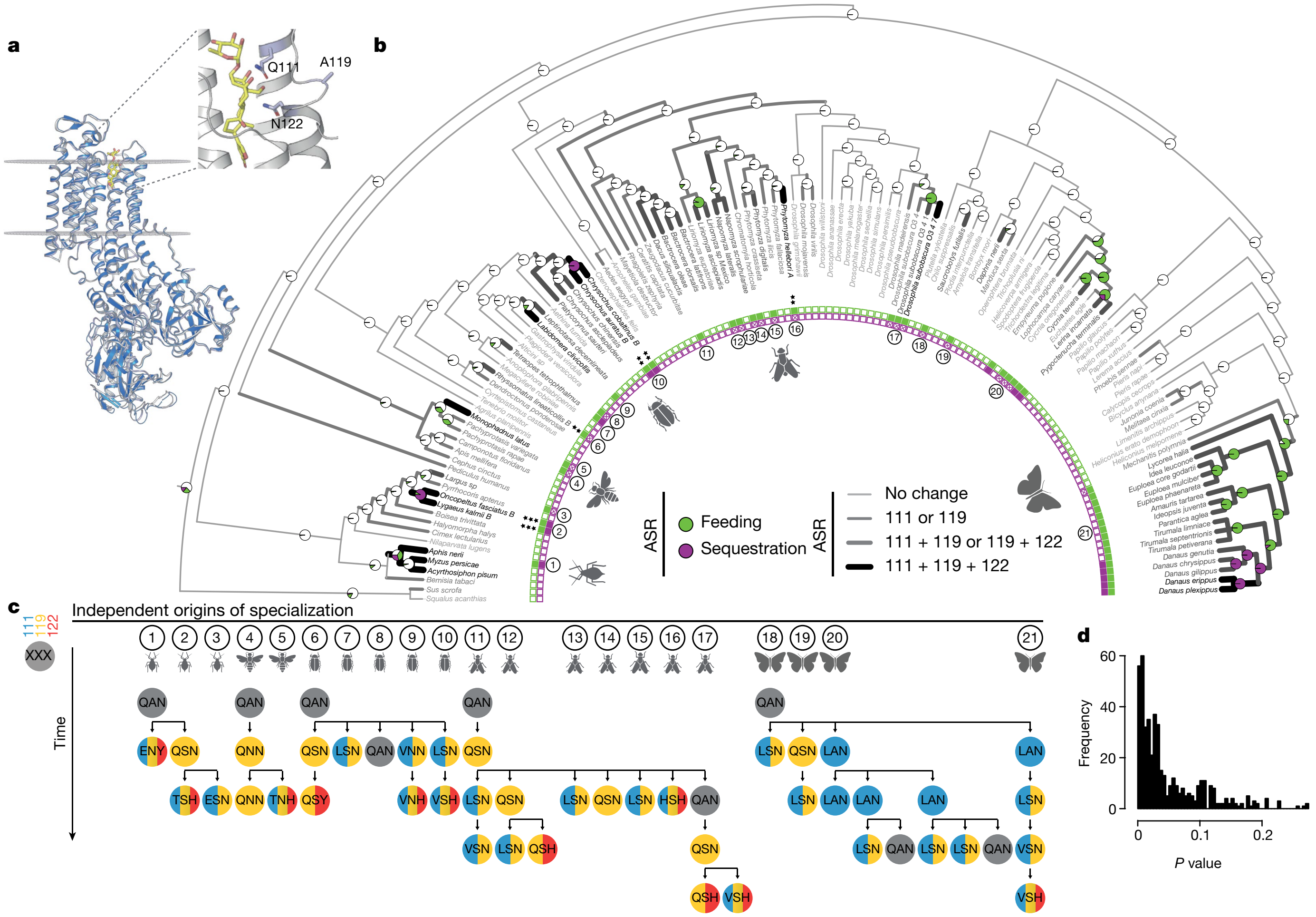


**Fig. 1 | Mutational paths in ATPα associated with insect specialization on cardiac glycoside-producing plants are constrained. a**, Protein homology model of *Drosophila melanogaster* ATPα (navy) superimposed on a *Sus scrofa* ATPα crystal structure (light grey) with ouabain (yellow) in the binding pocket. Residues 111, 119 and 122 (sticks) within the H1–H2 extracellular loop are associated with feeding on cardiac glycoside-producing plants and toxin sequestration. **b**, Maximum likelihood phylogeny based on 4,890 bp from *Atpα* and *coi*, with maximum likelihood ancestral state reconstruction (ASR) of feeding and sequestering states, estimated from the states of extant species (inner band of squares). Reconstructions are shown as nodal pie graphs (white, neither feeding nor sequestering; green, feeding; purple, feeding and sequestering), and the number of substituted sites at positions 111, 119 and 122 along branches in grey-scale (light grey 0, medium grey 1, dark grey 2, black 3), based on maximum likelihood ASR of H1–H2 loop amino acid sequences. Black asterisks indicate the *Atpα* copy number for species with multiple paralogues. **c**, ATPα substitutions inferred from ASR at positions 111 (blue), 119 (yellow) and 122 (red) in 21 lineages where specialization occurred independently. **d**, *P* value distribution from a set of randomized tests to determine the reproducibility of substitutions observed along mutational paths among sub-sampled groups compared to randomly permuted substitutions. On average, 4.9% (considering all mutational steps) of randomly permuted trajectories demonstrate a degree of ordering equal to or greater than observed mutational paths.

of ouabain, a hydrophilic cardiac glycoside[6] (Fig. 2c, d, Extended Data Figs. 4, 5). LAN, the first genotype to evolve, increased larval–adult survival at lower ouabain concentrations, but survival declined sharply as concentrations increased. LAN also increased adult survival at lower ouabain concentrations. LSN, the second genotype to evolve, increased larval–adult survival at the highest ouabain concentrations. The next step, VSN, provided the same larval–adult and adult survival benefit as LSN. Finally, the survival of 'monarch flies' carrying the monarch butterfly genotype (VSH) was unaffected by even the highest levels of ouabain in larvae and adults[6,9,11,18] (Fig. 2c, d), which was not due to reductions in feeding rate or toxin ingestion (Extended Data Fig. 6).

When knock-in line eggs were placed on medium containing the suite of cardiac glycosides found in the leaves of the milkweed species *Asclepias curassavica* and *A. fascicularis*[6], monarch lineage fly genotypes generally showed increased egg–pupal and egg–adult survival rates (Fig. 2e, Extended Data Fig. 7), although not always for VSN (Extended Data Figs. 3, 7). The LSN, VSN and VSH genotypes may enable insects to cope with the complex milieu of cardiac glycosides encountered during host shifts to these plants.

The monarch butterfly ATPα substitutions at positions 111, 119 and 122 may unlock a passive evolutionary route to cardiac glycoside sequestration, as we found small amounts of ouabain in newly emerged adult 'monarch flies' reared as larvae on a diet containing ouabain (Fig. 2f). However, toxin concentrations were far lower than in monarch butterflies, and the location of ouabain in flies is unclear[6,17,18].

At the physiological level, sodium pump enzymatic assays using head extracts from knock-in fly lines showed that each sequential monarch lineage genotype had a neutral-to-positive effect on TSI to ouabain (Fig. 3a). LAN provided a small increase in TSI, while the next genotypes to evolve, LSN and VSN, increased TSI by about ten times (Fig. 3a). Remarkably, TSI rose about a thousand times in mutant flies carrying VSH, the monarch butterfly genotype. 'Monarch fly' sodium pumps were as insensitive to ouabain as those of monarch butterflies (Fig. 3a). Sodium pump enzymatic assays using extracts of *Spodoptera frugiperda* Sf9 cells transiently expressing *Drosophila* ATPα proteins with key monarch lineage mutations at sites 111, 119, and 122 mirrored these patterns[24] (Extended Data Fig. 8). VSH is sufficient for 'monarch flies' to achieve the same degree of TSI to ouabain as the monarch

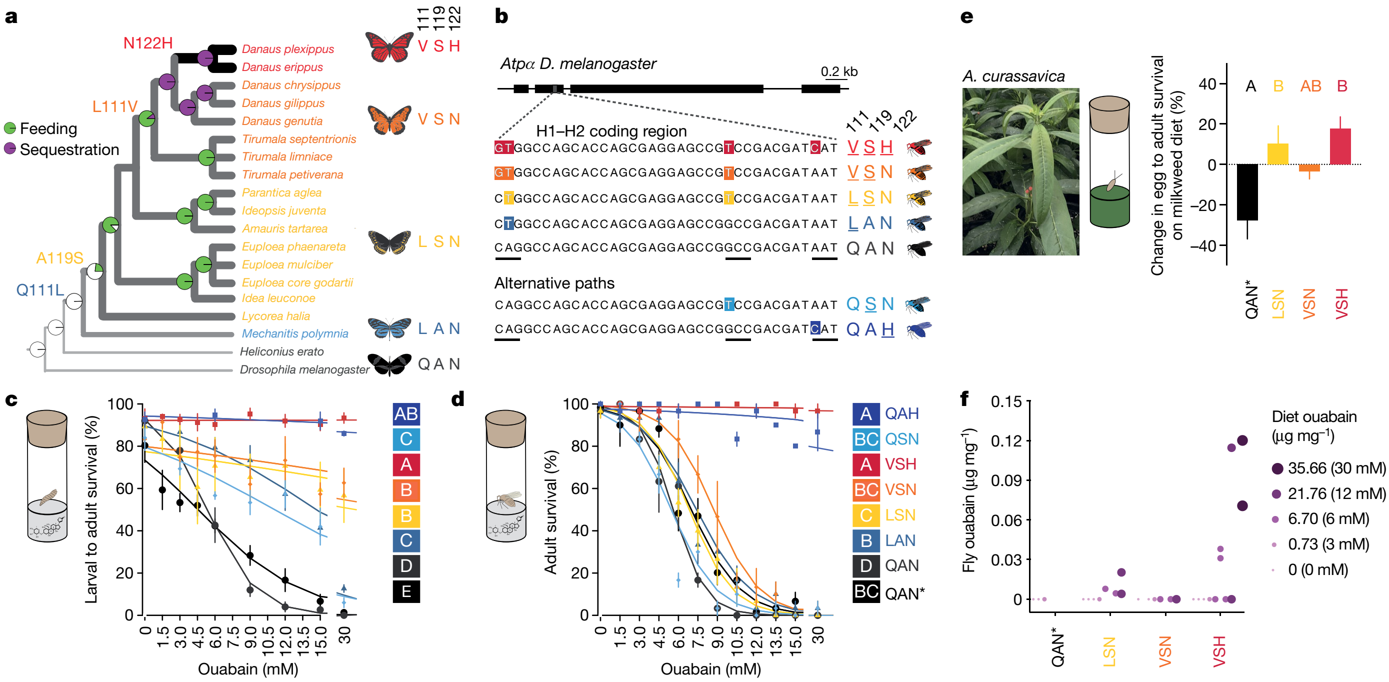


**Fig. 2 | *Drosophila* flies with edited genomes retrace the mutational path of ATPα in the monarch butterfly lineage and show increased survival when fed cardiac glycosides. a**, The monarch butterfly lineage with the substitutions observed in the H1–H2 loop of ATPα (adapted from Petschenka et al.)[11]. **b**, Non-synonymous point mutations in the edited DNA sequence of the native *Atpα* in *Drosophila* knock-in lines code for the substitutions at sites 111, 119 and 122. Codons are underlined. **c**, **d**, Larval–adult survival (**c**) and adult survival (**d**) of flies reared on diets with ouabain were different between monarch lineage knock-in lines and control lines (QAN = engineered control; QAN* = $w^{1118}$ wild type). Symbols represent the mean ± s.e.m. of 3–6 biological replicates (50 larvae and 10 females per replicate in **c** and **d**, respectively). Curves were fit using a logistic regression model for each line. Pairwise differences in survivorship trajectories between lines were evaluated with a likelihood ratio test on the significance of the interaction term between genotype (line) and ouabain concentration in a logistic regression for each pair of lines (letters). **e**, Egg–adult survival on diet supplemented with *Asclepias curassavica* leaves relative to control diets ($n = 3$–$4$; 100–200 eggs per replicate, see Methods; mean ± s.e.m.) was different between monarch lineage knock-in lines and QAN* (one-way ANOVA, $P = 0.0035$ followed by post hoc Tukey's tests (letters)). **f**, Ouabain concentrations in diet versus adult fly bodies among monarch lineage knock-in lines ($n = 2$–$4$ biological replicates per group). Adult flies had not fed since eclosion. Genotype and dietary ouabain concentration influenced the probability of detecting ouabain in post-eclosion flies (logistic regression and likelihood ratio test, genotype two-sided $P = 0.024$, dietary ouabain concentration two-sided $P = 6.344 \times 10^{-5}$). Further information on experimental design and statistical test results is found in the Source Data.

butterfly, suggesting that TSI is the predominant biological mechanism for the in vivo toxin resistance observed above.

Our results provide a starting point for identifying a biochemical mechanism for TSI through molecular docking simulations of ouabain bound to homology-modelled *Drosophila* $Na^+/K^+$-ATPase. In keeping with in vivo and enzymatic results (Figs. 2c–f, 3a), docking scores predicted that the first three *Atpα* genotypes to evolve resulted in slight increases in TSI to ouabain, followed by a larger increase for VSH, consistent with ouabain binding affinity measurements (Fig. 3b, Extended Data Fig. 9). Elucidation of the full biochemical mechanism for TSI will require additional research.

Site 119 co-evolves with site 111 (Extended Data Fig. 1c), and substitutions at site 119 always occurred before or with TSI-conferring substitutions at site 122 (Fig. 1c), suggesting that antagonistic pleiotropy and epistasis may have shaped mutational paths to resistance and TSI. We further addressed this hypothesis using knock-in lines for N122H (QAH), which is often the last substitution to evolve, and A119S (QSN), a substitution found in both specialists and non-specialists (Fig. 1b, c). QSN did not increase larval–adult survival at lower ouabain concentrations, but provided a survival benefit as ouabain concentrations rose, before survival dropped sharply at the highest concentration, as in LAN (Fig. 2c). In adults, QSN provided a slight survival benefit against ouabain (Fig. 2d), and low TSI to ouabain in sodium pump enzymatic assays (Fig. 3a). At lower ouabain concentrations, LSN provided an increase in larval–adult survival similar to that of LAN, but survival increased by about 40% over LAN and QSN at the highest concentration (Fig. 3c). This points to a role for synergistic, beneficial epistasis in dietary ouabain resistance between Q111L and A119S. QAH provided nearly as much survival benefit as VSH in larval–adult and adult feeding experiments (Fig. 2c, d), and in sodium pump enzymatic assays, TSI to ouabain in QAH was second only to VSH (Fig. 3a). Finally, the binding affinity of ouabain to $Na^+/K^+$-ATPase is high for QSN, and low for QAH, according to docking scores (Fig. 3b, Extended Data Fig. 9).

N122H conferred the highest TSI of any substitution, yet appeared last in the adaptive walk, and was contingent on a substitution at site 119 (Fig. 1c, d), which suggests that N122H imposes high fitness costs that are mitigated by A119S. To investigate this, we phenotyped monarch lineage knock-in flies for neurological seizures upon shaking (bang sensitivity), a common phenotype in hypomorphic $Na^+/K^+$-ATPase mutants[29,30]. Bang sensitivity varied widely within and among knock-in fly lines; this variation could be due to the intentionally introduced mutations or to unidentified (epi)mutations that arose from base editing and resulted in cryptic decanalization effects on the function of the nervous system. Despite this uncertainty, QSN flies were the least bang sensitive, and QAH flies were the most (Fig. 3d). Furthermore, the first step in the adaptive walk, LAN, resulted in higher bang sensitivity than the second step, LSN. Some costs of Q111L are therefore mitigated by A119S. When N122H was added to VSN, resulting in VSH (the monarch butterfly genotype), the cost of N122H was also reduced, in part, by A119S (Fig. 3d). A119S ameliorates the pleiotropic costs of resistance substitutions at sites 111 and 122 throughout the adaptive walk.

Substitutions at three amino acid sites in ATPα are sufficient together, but not alone, to explain the evolution of resistance and TSI to cardiac glycosides achieved by the monarch butterfly at organismal, physiological and biochemical levels. The adaptive walk follows theoretical predictions on the length of such walks[2–4,13,14], involves

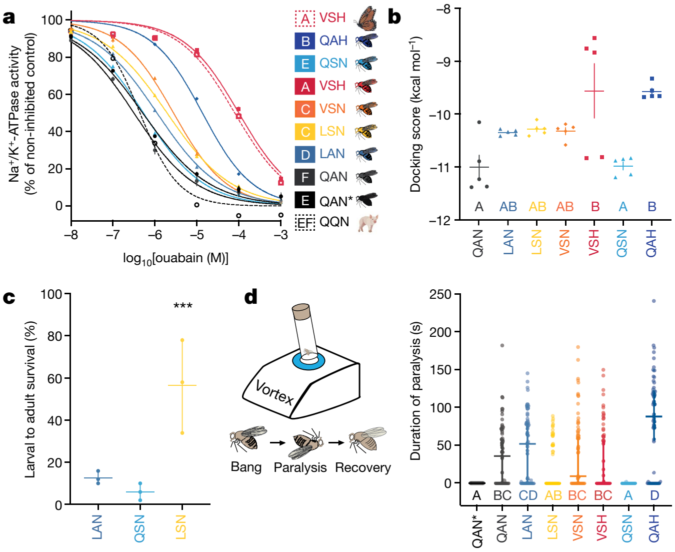


**Fig. 3 | The mutational path of ATPα in the monarch butterfly lineage sequentially increases TSI to ouabain and is shaped by epistasis.** **a**, In vitro ouabain sensitivity of $\mathrm{Na^+/K^+}$-ATPase activity in extracts of monarch lineage knock-in and control line fly heads (solid lines; QAN, engineered control; QAN*, $w^{1118}$ wild type), against activity in extracts of monarch butterfly and pig nervous tissue (positive and negative control, dashed red and black line, respectively). Symbols represent the mean ± s.e.m. of 3–7 biological replicates. $\log_{10}[\mathrm{IC}_{50}]$ (half-maximum inhibitory concentration) values for the $\mathrm{Na^+/K^+}$-ATPases were estimated after fitting four-parameter logistic regression curves, and were different between genotypes (one-way ANOVA ($P < 0.0001$) with post hoc Tukey's tests (letters)). **b**, Mean docking scores (± s.e.m. of five replicate calculations) from molecular simulations of ouabain binding to the $\mathrm{Na^+/K^+}$-ATPases found along the monarch lineage showed differences between genotypes (one-way ANOVA ($P = 0.0001$) with post hoc Tukey's tests (letters)). **c**, Effects of the substitutions Q111L, A119S and their combination on larval–adult survival on diets with 30 mM ouabain. Symbols represent the mean ± s.e.m. of three biological replicates (50 larvae each). The effect of mutations A119S and Q111L together was nearly threefold greater than the combined individual effects on survivorship (logistic regression, interaction effect between mutations: $^{***}P = 2.36 \times 10^{-15}$), indicating positive epistasis. **d**, Duration of paralysis following mechanical shocks (that is, bang sensitivity; $n = 60$ five-to-six-day-old adult flies). Bang sensitivity was affected by genotype (Kruskal–Wallis test ($P < 0.0001$) with post hoc Dunn's multiple comparisons tests (letters); medians with 95% confidence intervals), and was higher for QAH than for all other genotypes ($P < 0.05$), except for LAN, which showed higher bang sensitivity than LSN ($P = 0.0134$). Further information on experimental design and statistical test results can be found in the Source Data.

epistasis[13,14,20,22], and minimizes pleiotropic fitness costs[3,4,13,14,21], and variations of it convergently re-appeared across lineages that diverged more than three hundred million years ago[7–11]. Genome editing technology facilitates functional tests of adaptation across levels of biological organization[5,25,26]. Although mutational paths to adaptive peaks have been identified in microorganisms[2–4,13,14,22], this is, to our knowledge, the first in vivo validation of a multi-step adaptive walk in a multicellular organism, and illustrates how complex organismal traits can evolve by following simple rules.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1610-8.

Received: 26 October 2018; Accepted: 3 September 2019;

1. Barrett, R. D. & Hoekstra, H. E. Molecular spandrels: tests of adaptation at the genetic level. *Nat. Rev. Genet.* **12**, 767–780 (2011).
2. Agrawal, A. A. Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *Am. Nat.* **190**, S1–S12 (2017).
3. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
4. Storz, J. F. Causes of molecular convergence and parallelism in protein evolution. *Nat. Rev. Genet.* **17**, 239–250 (2016).
5. Siddiq, M. A., Loehlin, D. W., Montooth, K. L. & Thornton, J. W. Experimental test and refutation of a classic case of molecular adaptation in *Drosophila melanogaster*. *Nat. Ecol. Evol.* **1**, 0025 (2017).
6. Agrawal, A. A., Petschenka, G., Bingham, R. A., Weber, M. G. & Rasmann, S. Toxic cardenolides: chemical ecology and coevolution of specialized plant–herbivore interactions. *New Phytol.* **194**, 28–45 (2012).
7. Petschenka, G., Wagschal, V., von Tschirnhaus, M., Donath, A. & Dobler, S. Convergently evolved toxic secondary metabolites in plants drive the parallel molecular evolution of insect resistance. *Am. Nat.* **190**, S29–S43 (2017).
8. Holzinger, F., Frick, C. & Wink, M. Molecular basis for the insensitivity of the Monarch (*Danaus plexippus*) to cardiac glycosides. *FEBS Lett.* **314**, 477–480 (1992).
9. Dobler, S., Dalla, S., Wagschal, V. & Agrawal, A. A. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na,K-ATPase. *Proc. Natl Acad. Sci. USA* **109**, 13040–13045 (2012).
10. Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. & Andolfatto, P. Parallel molecular evolution in an herbivore community. *Science* **337**, 1634–1637 (2012).
11. Petschenka, G. et al. Stepwise evolution of resistance to toxic cardenolides via genetic substitutions in the $\mathrm{Na^+/K^+}$ -ATPase of milkweed butterflies (lepidoptera: Danaini). *Evolution* **67**, 2753–2761 (2013).
12. Horisberger, J. D. Recent insights into the structure and mechanism of the sodium pump. *Physiology (Bethesda)* **19**, 377–387 (2004).
13. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
14. Toprak, E. et al. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44**, 101–105 (2011).
15. Whiteman, N. K. & Mooney, K. A. Evolutionary biology: Insects converge on resistance. *Nature* **489**, 376–377 (2012).
16. Berenbaum, M. R. in *Molecular Aspects of Insect–Plant Associations* (eds Brattsten, L. B. & Ahmad, S.) 257–272 (Springer, 1986).
17. Brower, L. P., Ryerson, W. N., Coppinger, L. L. & Glazier, S. C. Ecological chemistry and the palatability spectrum. *Science* **161**, 1349–1350 (1968).
18. Petschenka, G. & Agrawal, A. A. Milkweed butterfly resistance to plant toxins is linked to sequestration, not coping with a toxic diet. *Proc. R. Soc. B* **282**, 20151865 (2015).
19. Zhan, S. et al. The genetics of monarch butterfly migration and warning colouration. *Nature* **514**, 317–321 (2014).
20. Tarvin, R. D. et al. Interacting amino acid replacements allow poison frogs to evolve epibatidine resistance. *Science* **357**, 1261–1266 (2017).
21. Hague, M. T. J. et al. Large-effect mutations generate trade-off between predatory and locomotor ability during arms race coevolution with deadly prey. *Evol. Lett.* **2**, 406–416 (2018).
22. Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489**, 513–518 (2012).
23. Groen, S. C. et al. Multidrug transporters and organic anion transporting polypeptides protect insects against the toxic effects of cardenolides. *Insect Biochem. Mol. Biol.* **81**, 51–61 (2017).
24. Dalla, S. & Dobler, S. Gene duplications circumvent trade-offs in enzyme function: Insect adaptation to toxic host plants. *Evolution* **70**, 2767–2777 (2016).
25. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
26. Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc. Natl Acad. Sci. USA* **111**, E2967–E2976 (2014).
27. Pegueroles, C. et al. Inversions and adaptation to the plant toxin ouabain shape DNA sequence variation within and between chromosomal inversions of *Drosophila subobscura*. *Sci. Rep.* **6**, 23754 (2016).
28. Shorrocks, B. in *The Genetics and Biology of Drosophila* (eds Ashburner, M. et al.) 385–428 (Academic, 1982).
29. Ashmore, L. J. et al. Novel mutations affecting the Na, K ATPase alpha model complex neurological diseases and implicate the sodium pump in increased longevity. *Hum. Genet.* **126**, 431–447 (2009).
30. Schubiger, M., Feng, Y., Fambrough, D. M. & Palka, J. A mutation of the *Drosophila* sodium pump α-subunit gene results in bang-sensitive paralysis. *Neuron* **12**, 373–381 (1994).

# METHODS

**Data reporting.** We performed preliminary experiments to assess variance and estimate sample sizes before conducting all bioassays. We used the online tool random.org (https://www.random.org/sequences/) to randomize all bioassays, except for the feeding experiments with milkweed and the sequestration assay. We were blind to the allocation of individuals to treatment groups. We were also blind to the treatment groups when taking measurements during all bioassays, except for the feeding experiments with milkweed and the sequestration assay.

**Species phylogeny and ancestral state reconstructions.** To identify amino acid changes associated with transitions to specialization on cardiac glycoside-producing host plants and cardiac glycoside sequestration, we first constructed a species maximum likelihood (ML) phylogeny. Phylogenetic relationships between all surveyed species were taken from previous studies (see Supplementary Table 1 for references). The phylogeny included insects adapted to feeding on such plants, their non-adapted relatives, and two vertebrate outgroup species. The species sampling method is described in the Supplementary Text. Because we constructed this supertree from disparate studies that used overlapping but non-identical molecular markers, we estimated branch lengths under this fixed topology using the full-length coding sequences of *Atpα* and the widely used gene coding for the mitochondrial cytochrome *c* oxidase subunit 1 (*coi*) for a combined total of 4,890 bp of sequence (see Supplementary Table 2 for ID numbers, source databases and references). Branch length estimation was implemented in IQ-TREE version 1.6.6[31]. The amino acid residues were numbered after the mature *Sus scrofa* protein residue numbering as in previous phylogenetic studies of *Atpα* in insects[7–11,24]. For each of the species with *Atpα* duplications, we included only the paralogue that is mostly expressed in the insect's body rather than copies that show expression bias towards the nervous tissue, so as to be conservative in downstream analyses[10]. Our reasoning is that the body-expressed copies experience stronger selective pressure from exposure to dietary cardiac glycosides than nervous tissue-expressed copies because the nervous tissue is protected from cardiac glycosides by the haemolymph–brain barrier (perineurium)[6,23,32]. Thus, we focused on the origin of substitutions that fixed as a result of interactions with cardiac glycosides, rather than adaptation to the nervous system after release from this agent of selection.

We performed ASR of feeding on cardiac glycoside-producing plants and cardiac glycoside sequestration by ML using the ace function of the ape package in R[33]. We also performed ASR of amino acid sequences of the H1–H2 extracellular loop using joint ML methods[34], implemented in HyPhy version 2.3.14[35]. Amino acid sites that were (1) identified as coevolving (see description of Bayesian graphical model (BGM) methods below), (2) significantly associated with cardiac glycoside feeding and sequestration (see TraitRateProp analysis below), and (3) highly variable and evolving convergently (see section Reproducibility of mutational order)—sites 111, 119, and 122—were chosen as the focus of this study. Using the ASR of amino acid sequences, the number of substitutions at these three sites was mapped onto the species tree, along with ancestral states of cardiac glycoside feeding and sequestration. The tree was visualized using the Interactive Tree Of Life (iTOL)[36], and can be found here: https://itol.embl.de/tree/1361521431149621561136381.

Using the joint likelihood framework of TraitRateProp[37], we identified amino acid residues within the H1–H2 loop of ATPα for which the rate of sequence evolution was associated with changes in the cardiac glycoside feeding and sequestering character state described above. More details of this analysis can be found in the Supplementary Text.

To detect epistatic or co-evolving interactions between amino acid sites within the H1–H2 loop that might constrain or enable the evolution of ATPα, we used the Spidermonkey/BGM method[38] in HyPhy version 2.3.14. Sites in this analysis were filtered with a default minimum number of five substitutions across the phylogeny. Interactions between sites with a marginal posterior probability (PP) $> 0.5$ were considered biologically meaningful.

*Convergent substitutions in the amino acid residues of the H1–H2 loop of Atpα.* We aimed to identify the most variable and convergently evolving amino acid residues within the H1–H2 loop in extant insects of 21 lineages in which specialization on cardiac glycoside-producing host plants and toxin sequestration evolved independently according to ASR analyses. At each position in the H1–H2 loop, we quantified the number of times a particular substitution occurred across these independent lineages (for example, the number of times Q111L occurred). We used the results from the ASR for feeding on cardiac glycoside-producing plants and sequestration to define these 21 independent lineages that evolved these traits, and we used the ASR for the H1–H2 loop to identify independent amino acid changes along these lineages (see above). We used the following set of rules to estimate the number of substitution occurrences across these lineages: (1) we considered substitutions only in extant species that were different from the ancestral sequence of all insect orders, (2) ancestrally shared substitutions in the H1–H2 loop were counted once to avoid overrepresentation, and (3) we did not score for reversion substitutions to the ancestral state. For further information on the estimation of independent substitution occurrences see the Source Data for Extended Data Fig. 1.

*Reproducibility of mutational order.* We identified amino acid residues 111, 119 and 122 of the H1–H2 loop of cardiac glycoside-adapted insects as the most variable and convergently evolving (to the same amino acid states). Next, we used a statistical method that compared the observed evolutionary pattern of mutations at ATPα positions 111, 119 and 122 in specialized insects to a random permutation null model to test whether the ordered appearance of these mutations could have occurred by chance[14]. The reproducibility of mutational order (RMO) score was defined as previously described[14]. The RMO score of two ordered sequences of ATPα mutations along insect lineages is defined by the number of shared mutation pairs that occurred in the same order, from which we subtract the number of shared mutation pairs that occurred in reverse order—for example, two sequences that experienced mutations among six sites, [1, 2, 3, 4] and [1, 6, 4, 3], would have an RMO score of 1 (two shared mutation pairs occurred in the same order [1 and 3; 1 and 4], but one shared mutation pair occurred in the opposite order [3 and 4]). We used the results from the ASR of feeding and sequestering in conjunction with the ASR results of the H1–H2 loop sequence to define the mutational path in each lineage examined, starting from an ancestor that was estimated to be neither feeding on cardiac glycoside-producing plants nor sequestering the toxins and ending at extant species that feed and/or sequester.

The RMO score for a set of two paths requires two assumptions that are violated in this study: that the exact order in which mutations arose can be discerned, and that each mutational path is independent[14]. Multiple mutations that co-occurred along the same branch leave the order in which each mutation occurred ambiguous. In addition, different paths can be dependent on one another as mutations arising on internal branches in the phylogeny can appear in multiple different paths towards branch tips that derived from these internal branches. Under this dependency, we find 13 groups of independent paths. To evaluate the probability of seeing these paths by chance, conditional on these two complicating factors, we used a two-step randomization process. First, we randomly sampled one path from each of the 13 groups of paths, for a total of 13 independent paths (see Source Data for Fig. 1d). Second, for any co-occurring mutations, we randomly chose a sequential ordering of them. We then calculated the RMO score for this set of sampled paths. Next, we randomly permuted the ordering of the mutations within each path, and calculated the RMO score of this random order. This random permutation was done 100,000 times to get 100,000 different RMO scores. We calculated the fraction of these RMO scores that were greater than or equal to the RMO score of the 13 paths originally sampled. The resulting *P* value represents the probability that the observed level of ordering in the 13 paths could have been produced by random chance, with a smaller *P* value indicating stronger evidence for non-random ordering of the observed mutations. This two-step randomization process was repeated 500 times, to generate a distribution of 500 unique *P* values. Through this process, we found a mean *P* value of 0.0495, indicating that on average less than 5% of the random 100,000 permutations induced an ordering of the same magnitude as our observed data. We also observed that 66.6% of the 500 *P* values generated were lower than 0.05. The heavy tail in large *P* values (the other 33.3%) can be attributed to the single path in which mutations at sites 111, 119, and 122 all occur on a single branch, and the ordering is left ambiguous. When we randomized the ordering of these three mutations, we found a mean *P* value of 0.0149 when the mutation at position 119 is more ancient than that at 122 on this specific path. Conversely, when the randomization chooses 122 to be more ancient than 119, we find a much larger mean *P* value of 0.073.

**Generation of knock-in fly lines.** *Genomic engineering strategy.* We used a two-step genome editing approach using CRISPR–Cas9 coupled with HDR to introduce the different non-synonymous point mutation combinations of interest into the native *Atpα* of *D. melanogaster* in the region coding for the enzyme subunit's H1–H2 extracellular loop (Extended Data Fig. 2). In the first step, the *Atpα* region encoding the H1–H2 loop was replaced with a 3×P3::GFP marker through CRISPR–Cas9-mediated HDR. This generated the deletion allele *Atpα*$^{\mathrm{Deletion\ (GFP+)}}$. This line served as a common stock for the second step, in which the 3×P3::GFP marker was replaced with each of the point mutation alleles in *Atpα* through an additional round of CRISPR–Cas9-mediated HDR. This generated a first set of knock-in lines with the non-synonymous point mutation alleles *Atpα*$^{\mathrm{Q111L/A119S}}$ (LSN), *Atpα*$^{\mathrm{Q111V/A119S}}$ (VSN), and *Atpα*$^{\mathrm{Q111V/A119S/N122H}}$ (VSH), and later, a second set of knock-in lines with the non-synonymous point mutation alleles *Atpα*$^{\mathrm{Q111L}}$ (LAN), *Atpα*$^{\mathrm{A119S}}$ (QSN), and *Atpα*$^{\mathrm{N122H}}$ (QAH). With this second set, we also generated a CRISPR–Cas9-engineered control line following the same two-step strategy but without introducing any non-synonymous changes (QAN; Supplementary Table 6).

*Construction of guide RNAs and donor plasmids.* All guide RNAs (gRNAs) used in the first and second rounds of CRISPR–Cas9-mediated HDR were cloned into the pCFD3-dU6:3gRNA vector[26], a gift from S. Bullock (Addgene plasmid 49410). All gRNAs were tested for potential off-target effects using the flyCRISPR website[39] (http://flycrispr.molbio.wisc.edu/). No potential off-target sites elsewhere in the genome were predicted, but they cannot be ruled out. The donor plasmid for the generation of the deletion line was created through Golden Gate assembly via BsaI digestion and subsequent ligation of four DNA fragments: a selection cassette containing the 3×P3::GFP marker (GenetiVision), two ~1-kb-sized, PCR-amplified homologous arms (by Phusion polymerase) and a pUC57-Kan vector backbone. Each of the donor plasmids for the generation of the mutant alleles (replacement vectors) was also created through Golden Gate assembly via BsaI digestion and subsequent ligation of four DNA fragments: the donor DNA containing the mutations of interest, two ~1-kb-sized, PCR-amplified homologous arms (by Phusion polymerase) and a pUC57-Kan vector backbone. For the replacement vectors, an initial donor plasmid was built that contained two point mutations at the gRNA PAM sequences to prevent CRISPR–Cas9 retargeting. Because the mutation in the region of the downstream gRNA codes for amino acid, a silent mutation at E176 (GAG to GAA) was included in the design as a control. This initial donor plasmid served as a template for introducing the point mutations in the donor DNAs for all alleles through template extension PCR. The sequences of the gRNAs and recovered *Atpα* in all homozygous knock-in lines are provided in Supplementary Tables 4 and 6, respectively.

*Embryo injection, crosses and mutant screening.* For the generation of the *Atpα* deletion line, a mixture of two gRNAs (at 50 ng/µl each) and the donor plasmid carrying the 3×P3::GFP marker (at 150 ng/µl) was injected into embryos of BL#54591 carrying the *nos-Cas9* gene[26]. The adult flies that survived microinjection (G0) were crossed with $w^{1118}$ and G1s were screened for the presence of green fluorescence in the compound eyes. Positive male heterozygous G1s were selected and genotyped by PCR with cassette-specific primers and Sanger sequencing (in both directions) to confirm that the cassette insertion was in the expected location. For this, two sets of PCR primers were used that covered both the left homologous arm, plus its junction upstream of the GFP cassette, and the right arm, plus its junction downstream of the cassette. The positive G1s were then crossed to the double-balanced *TM3, Sb/TM6, ebony* fly line and screened for the absence of the *ebony* phenotype. Positive G2 ATP$\alpha^{\text{Deletion (GFP+)}}$/*TM3*, *Sb* flies were obtained and the *Atpα* deletion line was established.

For the generation of each *Atpα* knock-in line, a mixture of two gRNAs (at 50 ng/µl each) and each donor plasmid carrying the respective point mutation(s) (at 150 ng/µl) was injected into embryos of the *Atpα* deletion line carrying *nos-Cas9*. To achieve this, the *Atpα* deletion line was crossed with BL#54591 again to introduce *nos-Cas9* into the X chromosome of the deletion line. The injected G0 flies were crossed with the double-balanced *TM3*, *Sb*/*TM6*, *ebony* fly line and screened for the absence of green fluorescence in the compound eyes. Positive male heterozygous G1s were selected, genotyped by PCR with primers covering the region with the knock-in mutations, and balanced with *TM3*, *Sb*. Positive heterozygous G2 flies were intercrossed and G3 homozygous flies were obtained carrying the desired knock-in mutations. These flies were genotyped with primers flanking the area affected by the genome engineering process at both steps of HDR. The overall crossing scheme for the generation of the *Atpα* knock-in lines is presented in Extended Data Fig. 2.

The sequences of all primers used for screening and validating the mutant lines are provided in Supplementary Table 5. The validated sequences of the knock-in and control lines are provided in Supplementary Table 6. The number of embryos injected, the number of independent lines established, and the ability of mutant lines to give rise to homozygous progeny are provided in Supplementary Table 7. GenetiVision performed all injections as well as the initial mutant screening.

**Egg lethality assay.** To compare egg–adult survival among the knock-in lines and the $w^{1118}$ (QAN*) line, we introduced 100 eggs of each line into narrow vials containing ~10 ml of 22% Instant *Drosophila* Medium in Millipore water (w/v). Each of the groups was tested in 7 or 8 replicate vials. Pupariation, adult eclosion, and survival were monitored over a period of ~21 days until all adult flies had eclosed.

**Real-time quantitative PCR.** Total RNA was extracted from five- or six-day-old females using the ReliaPrep RNA Tissue MiniPrep System including the DNase treatment step. From this, 900 ng total RNA was converted to cDNA using the ProtoScript II cDNA synthesis kit (NEB). Real-time quantitative PCR (qPCR) reactions were run on the StepOne Real-Time PCR System (ThermoFisher Scientific). Reaction volumes were as follows: 10 µl 2× DyNAmo HS SYBR Green qPCR Kit, 0.15 µl ROX Passive Reference Dye, 0.5 µl 60 µM forward and reverse primers, and 20 ng cDNA to a total reaction volume of 20 µl. qPCR was performed for *Atpα* using *rpl32* as an internal control. The primer sequences are provided in Supplementary Table 8. The *rpl32* primer sequences were as previously reported[40]. All runs included an initial 10-min denaturation step at 95 °C, followed by 40 cycles of 95 °C for 15 s and 58 °C for 1 min. The runs were finished with a melt curve ramp from 60 °C to 95 °C during which data were collected every +3 °C. The expression of *Atpα* for each genotype was assayed in two or three biological replicates and two technical replicates.

**In vitro sodium pump assays using extracts of heads and nervous tissue.** *Sodium pump preparations.* Sodium pump preparations were performed as described[11]. For each genotype, 100 frozen flies (50 females and 50 males that were eight days old) were decapitated under a dissecting microscope, on dry ice, using a clean scalpel. Heads were then suspended in 2.2 ml Millipore water, homogenized (Wheaton glass grinder) to yield a single stock preparation, and pipetted into 250-µl aliquots (each aliquot contained the equivalent of 12.5 fly heads). Similarly, a pooled stock of monarch butterfly (*Danaus plexippus*) nervous tissue homogenate was prepared from adult butterflies that had been frozen alive at −80 °C and thawed on ice for dissection. In brief, 16 butterfly brains (with thoracic ganglia) were homogenized as above, pooled in 8 ml Millipore water, and pipetted into 375-µl aliquots (each aliquot contained the equivalent of three-quarters of a brain). All aliquots were frozen at −80 °C, freeze-dried overnight, and then stored at −80 °C until use. Purified porcine sodium pump (A7510, Sigma) was stored at −80 °C, at a concentration of 1 unit/ml in Millipore water, in 50-µl aliquots.

Monarch butterfly and fly sodium pump preparations were thawed on ice, resuspended in cold Millipore water (fly preparations: 425 µl, monarch butterfly preparation: 2 ml; this equates to 12.5 fly heads per 425 µl and 75 butterfly heads per 2 ml), vortexed, sonicated (ultra-sonic bath), and centrifuged for 5 min at 5,000 r.p.m. to remove pelleted material. Porcine sodium pump preparation was thawed on ice and diluted with deionized water to yield a concentration of 0.05 units/ml. All enzyme preparations were kept on ice during experiments.

*Plate assays.* For each enzyme type, we analysed activity against a serial dilution of ouabain (ouabain octahydrate, O3125, Sigma) ranging from $10^{-3}$ M to $10^{-8}$ M, in 10% DMSO. Three or four technical replicates of each fly preparation were assayed across three microplates, such that each plate included at least one replicate of each fly line, as well as monarch and pig sodium pump. Reactions were performed as described[11], except for the following change: instead of pre-incubation without ATP, we prepared a 'mastermix' for each enzyme type, on ice. This mastermix contained all reagents except for the inhibitor, and was dispensed directly onto the ouabain solution in the microplate just before incubation. After 20 min, reactions were stopped and inorganic phosphate was stained.

*Curve fitting and estimation of $IC_{50}$.* The absorbance values of reactions were corrected for the background value (full inhibition) for each enzyme type and calculated as per cent residual activity using reference points of the non-inhibited reaction as 100% and the completely inhibited reaction as 0% residual activity. The results for each enzyme type were fitted to the four-parameter logistic equation log(inhibitor) versus response–variable slope: $Y = \text{Bottom} + (\text{Top} - \text{Bottom})/1 + 10^{((\log_{10}[\text{IC}_{50}] - X) \times \text{Hillslope})}$ where $Y$ is the percentage of non-inhibited control; Bottom is the percentage of non-inhibited control (constrained to 0 as 0% residual activity of the reaction is expected in high doses of ouabain); Top is the percentage of non-inhibited control (constrained to 100 as 100% residual activity of reaction is expected in the absence of ouabain); $\text{IC}_{50}$ is the concentration of ouabain at which 50% of sodium pump activity is inhibited; $X$ is the concentration of ouabain; and Hillslope is the steepness of the curve.

Finally, to calculate the absolute activity of the $\text{Na}^+/\text{K}^+$-ATPase, a phosphate calibration curve ranging from 0.2 to 1.2 mM $\text{KH}_2\text{PO}_4$ was run on each plate, ouabain inhibition of the samples was analysed, and phosphate released (nmol/µl) was determined at 700 nm according to a previously described photometric method[41]. The activity of the $\text{Na}^+/\text{K}^+$-ATPase of each sample can be determined as the difference in the amount of phosphate released between the non-inhibited control (all ATPases active) and a reaction in which only the contaminating ATPases are active (a buffer system without KCl[42]). As the reactions were performed per fly head, the final $\text{Na}^+/\text{K}^+$-ATPase activity is expressed as nmol ATP hydrolysed per fly head per min.

**Feeding experiments with knock-in fly lines.** We performed all feeding experiments in a growth chamber at ~21–23 °C and ~50–60% humidity, on a 12 h:12 h light:dark cycle. In all feeding experiments, vials for independent trials were coded and placed in a randomized order in rows on cardboard trays.

*Feeding experiment with knock-in adults and larvae reared on medium containing ouabain.* For the immatures, we introduced fifty second-instar larvae in wide vials containing 10 ml of 27.5% Ward's Instant *Drosophila* Medium (470024-740, Ward's Science) in Millipore water (w/v) supplemented with a series of ouabain concentrations. Ouabain (≥95% purity) was obtained from Sigma-Aldrich (11018-89-6). We used second-instar larvae for the experiment because of the occurrence of variably penetrant embryonic lethality in the knock-in lines (Extended Data Fig. 3a). We selected the Instant *Drosophila* Medium diet because our knock-in lines and the $w^{1118}$ flies (QAN*) performed well on it. We assessed larval performance for all genotypes on diets that ranged in ouabain concentration from 0 mM to 30 mM. A concentration higher than 30 mM could not be prepared because ouabain crystallized in water above this concentration. Each of the treatment and control groups was tested in triplicate vials. Pupariation, adult eclosion, and survival were monitored over a period of ~21 days until all adult flies had eclosed.

For the adults, we introduced ten females (4–7 days old) in wide vials containing 5 ml of 5% sucrose/1% agar in Millipore water supplemented with a series of ouabain concentrations. We performed experiments with females because they have been reported to be more sensitive to ouabain than males when ouabain is ingested orally at (near-)millimolar concentrations[23,43]. We reared flies on medium with ouabain concentrations ranging from 0 to 30 mM for all genotypes. Each treatment and control group was tested in triplicate vials. Survival of females was monitored over a 3–6-day period in each vial.

*CAFE assays with knock-in adults.* We measured the effects of different dietary ouabain concentrations on fly survival and food consumption rate using the CApillary FEeder (CAFE) assay as described[23], with small modifications. We also estimated the lethal ouabain dose necessary to kill 50% of the flies in each treatment group ($LD_{50}$). We introduced three females (four to seven days old) in each vial. Flies were supplied daily with two capillary tubes (5-μl calibrated pipets, 53432-706, VWR) containing 5% sucrose in Millipore water (w/v) supplemented with concentrations of ouabain that ranged from 0 to 30 mM for all genotypes. The daily replacement of the capillaries ensured the ad libitum availability of food. Each genotype was tested with five replicate vials for all treatment and control groups. The experiment lasted for seven days. Consumption was measured as described[44].

To estimate $LD_{50}$ we measured the cumulative amount of food consumed (in μl per fly) over the average number of days it took for the second out of three flies to die in each replicate vial. The average number of days was rounded to the nearest whole number of days. As no more than half of the flies died in some of the treatment groups, we proportionally extrapolated the amount of food it took to cause the proportion of deaths we observed in such a treatment group to the amount of food it would then take to cause the death of half of the flies in that treatment group. The daily consumption rate per fly (in μl per fly per day) was calculated over the number of days for which at least one fly was alive in each replicate vial. Consumption data from the day(s) leading up to the death of the last fly in a vial were not considered for rate calculation if the last living fly was no longer mobile. Survival was expressed as the treatment group average of the median number of days of fly survival per vial during the experiment.

**Fly survival on medium supplemented with milkweed leaf powder.** We performed two experiments in which we raised three knock-in (LSN, VSN and VSH) and $w^{1118}$ control (QAN*) line flies on *Drosophila* medium with and without pulverized leaves of two milkweed species (*A. fascicularis* or *A. curassavica*). The purpose of these experiments was twofold: (1) to compare the survival of the knock-in lines to the control line in fly medium; and (2) to compare the effects of milkweed leaves on survival of the knock-in and control lines. The experimental procedure was the same for both experiments. We introduced eggs of each line in narrow vials containing ~10 ml of 22% Instant *Drosophila* Medium in Millipore water (w/v) with or without 8% dried milkweed leaf powder. We added milkweed leaf powder to the fly diet at an amount similar to that added in the semi-artificial monarch butterfly diet[45]. We dried the milkweed leaves of each species in a drying oven at ~45 °C with airflow for ~48 h and then ground them to powder with a mortar and pestle. We collected leaves of *A. fascicularis* growing naturally outside the Berkeley Oxford Track Greenhouses (Berkeley, CA) in January 2018. We grew *A. curassavica* plants in the greenhouse of the VLSB building at UC Berkeley and collected leaves in April 2018.

We introduced 100 eggs into each vial for knock-in lines LSN and VSH, and the $w^{1118}$ control (QAN*) line. We introduced 200 eggs for knock-in line VSN into each vial because of the higher embryonic lethality in this line (Extended Data Fig. 3a). Each of the treatment and control groups was tested in quadruplicate vials. Pupariation, adult eclosion, and survival were monitored over a period of ~24 days until all adult flies had eclosed.

**Sequestration assays.** *Collection of fly samples.* To assess sequestration of ouabain in flies, we examined extracts of freshly eclosed adult females and males of each knock-in (LSN, VSN, VSH) and $w^{1118}$ control (QAN*) line raised as larvae in white fly medium (Ward's Instant *Drosophila* Medium) supplemented with a series of ouabain concentrations (0 mM, 3 mM, 6 mM, 12 mM and 30 mM). We let 25 female flies (~five to eight days old) lay eggs for four days in wide vials containing a diet with each respective concentration of ouabain in white medium on which larvae fed after eclosion. When larvae completed development, we transferred puparia into fresh, empty food vials on humidified filter paper for each condition to prevent freshly emerging adult flies from feeding and subsequently excreting ouabain before collection. We then collected newly eclosed adult flies within 2–3 h of emergence. The flies were stored immediately in groups of 25 in 1.5-ml collection tubes at −20 °C.

*Preparation of fly samples for high-performance liquid chromatography (HPLC).* Frozen flies were lyophilized and each set of 25 individuals was weighed on a microbalance. Samples were pooled, within fly line and diet type, into sets of 50–100 flies and transferred to 2-ml screw-cap tubes containing 0.9 g zirconia/silica beads (2.3 mm; Biospec). Samples were then extracted in 1 ml methanol using a FastPrep homogenizer (MP Biomedicals; three 45-s sessions at 6.5 m/s), and centrifuged for 12 min at 14,000 r.p.m. Then, 750 μl of cleared supernatant was transferred to a fresh tube and dried to completion in a vacuum concentrator (Labconco). Samples were then filtered (0.2 μm; Millipore) and 15 μl extract was injected on an Agilent 1100 series HPLC unit (Agilent). Compounds were separated on a Gemini C18 reversed-phase column (3 μm, 150 × 4.6 mm, Phenomenex). Cardiac glycosides were eluted on a constant flow of 0.7 ml/min with an acetonitrile–water gradient as follows: 0–8 min: 16% acetonitrile; 25 min: 70% acetonitrile; 30–40 min: 95% acetonitrile. Peaks were detected by diode array at 218 nm, and absorbance spectra were recorded from 200 to 300 nm. The ouabain peak was identified in sample extracts by comparing its retention time to an external ouabain standard (O3125, Sigma), and confirmed by observing a symmetrical spectral absorbance peaking at 218 nm[46]. Sample ouabain concentrations were then estimated relative to the known concentration of the external standard.

*Fitting HPLC data to a multiple logistic regression model.* The HPLC data for the knock-in lines were fitted to a multiple logistic regression model in R (dependent variable = ouabain detected or not): $Y = \text{intercept} + X$ where $Y$ is the presence of ouabain in the fly; intercept is fixed to $X = 0$ and $Y = 0$ (because in the absence of ouabain in the food there is no ouabain in the fly); and $X$ is the mean amount of ouabain in the diet (μg/mg). The $w^{1118}$ control (QAN*) line was omitted from this analysis because this line did not survive well enough on diets with millimolar levels of ouabain. In all cases, the diet's intended ouabain concentrations were used as a predictor variable rather than the measured concentrations in the diet. A likelihood ratio test in R was used to test for significance of each model term (genotype and diet).

**In vitro sodium pump assays using extracts of transiently transfected Sf9 cells.** Genetically engineered *D. melanogaster* sodium pumps expressed in Sf9 cells were obtained as described[24]. The enzymes combined the β-subunit Nrv2.2 (accession no. NM_001273235) with the α-subunit ATPα (accession no. HE962487) into which the substitutions Q111L and A119S had been introduced either singly or in combination by site-directed mutagenesis (QuickChange II XL Kit; Agilent Technologies). A triple-mutated construct coding for the Q111V, A119S, and N122H substitutions was also generated. The constructs were expressed by baculovirus infection of Sf9 cells (Bac-to-Bac Baculovirus Expression System, Thermo Fisher Scientific), and membranes of infected Sf9 cells were collected according to previously established protocols[24]. Extracted membrane proteins were subjected to sodium pump enzyme assays as in the assays with extracts of heads and nervous tissue above, the only differences being that assays were started by the addition of ATP as described[11], and that absolute activities of the $Na^+/K^+$-ATPase were expressed as nmol ATP hydrolysed per mg protein per min after calculation[24].

*Curve fitting and estimation of $IC_{50}$.* Curve fitting was performed as for the data on sodium pump activity from the assays with extracts of heads and nervous tissue.

**Protein homology modelling and in silico mutagenesis.** The structure of the *D. melanogaster* $Na^+/K^+$-ATPase is not available in the protein data bank (PDB) and had to be obtained via multi-template homology modelling[47]. We started by selecting the crystal structures of high-affinity $Na^+/K^+$-ATPase from *Sus scrofa* (PDB ID: 4HYT)[48] and *Bos taurus* (PDB ID: 4XE5)[49] kidney as templates. The $Na^+/K^+$-ATPase α-subunits of these species share 75.9%, and 76.1% sequence identities with the *D. melanogaster* ATPase α-subunit, respectively. We then performed homology modelling of the structure via Modeller[47] and PyMod 2[50], using a PyMOL[51] plugin to prepare, align and perform the modelling calculations. Homologous structures for the query sequence were identified through the BLAST search tool, after which the template and target sequences were aligned using ClustalW. For more accurate modelling the existing ligands were removed from the templates, and the N-terminal excess amino acids were taken out of the target sequence. The modelled structure was minimized using 80 steps of Steepest Descent, 20 steps of Conjugate Gradients, 20 steps of Quasi Newton, and 100 steps of Molecular Dynamics at room temperature.

**Molecular docking.** *Protein and ligand preparation for molecular docking.* We extracted the ouabain structure from the co-crystal structure of high-affinity $Na^+/K^+$-ATPase from *S. scrofa* (PDB ID: 4HYT)[48], and prepared the ligand (ouabain) and protein files for docking simulations using the AutoDock Tools (ADT) package[52] of MGLTools 1.5.6[53]. During docking calculations the protein structure was kept rigid and the ligand flexible, while all water molecules were removed and only polar hydrogen atoms were added to the modelled protein structure used. We assigned Gasteiger charges to the ligand, and saved the ligand and protein structures in PDBQT format for docking simulations.

*Molecular docking.* For docking calculations of the ouabain ligand into the modelled ATPase structure we used Autodock4 (version 4.2)[52]. Grid maps were prepared with the AutoGrid tool of the ADT package, resulting in a grid box of 60 × 60 × 80 Å, centred around the coordinates of ouabain in the co-crystal structure with 0.375 Å of spacing between grid points. We ran a Lamarckian genetic algorithm (LGA) to search for the best conformers during the calculations, keeping the docked ouabain coordinates closest to the coordinates of the ouabain ligand in the co-crystal structure as best conformer for each set of docking calculations. Autodock4 was run at default parameter settings for all docking simulations with the exception of the number of GA runs, which was set to 100. Repeating docking calculations five times for each system allowed for statistical analysis of the results. Each docked structure was inspected visually and all of the structure representations were prepared in PyMOL. A lower binding affinity in the output of the docking simulations corresponds to a better docking of the ligand to the receptor.

**Bang sensitivity assays.** We performed paralysis assays with five- or six-day-old females. Females were kept with males in standard molasses fly medium (UC Berkeley Biosciences Divisional Services, Fly Food Facility) until the day of the experiment. In the paralysis assay trials, vials were coded and placed in a randomized order in rows on cardboard trays. Independent experiments were performed across each cohort as described[54], but with small modifications. Individual, female adults of each genotype were transferred to empty, clean, culture vials 10 min before the start of the assay. Each vial with one female was vortexed at maximum speed (3,000 r.p.m.) for 10 s using a standard laboratory vortexer (Analogue vortex mixer, VWR), after which the time until recovery from paralysis was measured. The time until recovery is defined as the time after vortexing that is required for the fly to regain the ability to stand upright. Data from three independent runs of the experiment were pooled for each line and medians were compared using the Kruskal–Wallis tests with Dunn's post hoc tests corrected for multiple comparisons.

**Statistical analysis.** For all bioassays, we performed pilot experiments to estimate sample sizes. For the feeding experiments, the concentration ranges for ouabain were determined using pilot experiments. We used the Prism software package (GraphPad Prism 8) to plot and statistically analyse all data, except for the data described in Figs. 1d, 2c, d, f, 3c. For the RMO analysis in Fig. 1d we used a package in R that was based on code developed originally for MATLAB[14]. For Fig. 2c, d, to determine whether the genotype of fly lines had an effect on survival across ouabain concentrations, we tested for each pair of knock-in lines whether the two predicted survival curves were significantly different from one another. We first performed a logistic regression on the probability of survival against model terms for genotype and ouabain concentration, as well as an interaction term for the effects of genotype and ouabain concentration combined. For two lines, L1 and L2, the logistic regression is of the form $\log(P(\text{survival})/1 - P(\text{survival})) = B_0 + B_1\text{ouabain} + B_2 I(\text{L2}) + B_3(\text{ouabain} \times I(\text{L2}))$, where $I(\text{L2})$ is the indicator function for the genotype of line L2, and ouabain represents the concentration in mM. We then used a likelihood ratio test (LRT, d.f. $= 1$), to evaluate the hypothesis that the coefficient of the interaction term is different from 0, with a non-zero coefficient indicating that the genotypes of the two lines have significantly different effects on survival at different ouabain concentrations ($B_2$ captures the general effect on survival, independent of ouabain concentration). This LRT provided a $P$ value which we compared to a Bonferroni-corrected significance level of $\alpha = 0.05/28 = 0.0018$, where 28 was the number of pairwise tests performed between all knock-in and control lines.

For the ouabain sequestration experiment in Fig. 2f we used an LRT to evaluate the significance of each model term in the logistic regressions. For Fig. 3c, we tested for potential epistasis between the Q111L and A119S mutations for larval–adult survivorship at the maximum ouabain concentration of 30 mM by performing a logistic regression with interaction term: $\log(P(\text{survival})/1 - P(\text{survival})) = B_0 + B_S I(\text{S}) + B_L I(\text{L}) + B_{SL} I(\text{S}) I(\text{L})$, where $I(\text{S})$ and $I(\text{L})$ are the indicator functions for the presence of the A119S and Q111L mutations, respectively. We tested the hypothesis that the effect of mutations A119S and Q111L together, $B_{SL}$, was equal to the sum of each individual effect on survivorship (if $B_{SL} = B_S + B_L$). In all other experiments, we tested for differences between genotypes with either one-way or two-way ANOVA with Tukey's post hoc tests corrected for multiple comparisons, or the non-parametric Kruskal–Wallis test with Dunn's post hoc tests corrected for multiple comparisons, as appropriate. In experiments where four-parameter logistic equations were fitted to the data (Fig. 3a, Extended Data Fig. 8), we used global fitting followed by the extra sum-of-squared $F$ test as a method to test whether the model parameters differed between data sets. We provide the full statistical test results in the Source Data files for each figure.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The data supporting the findings of this study are available within the paper and its Supplementary Information.

## Code availability

The code to compute RMO (reproducibility of mutational order) scores in the presence of unordered mutations and correlated pathways can be accessed in Github (https://github.com/gaguerra/ModifiedRMO). The set of R scripts implements the RMO score, first proposed by Toprak and co-workers[14], with the new additions of accounting for non-independent mutational pathways (in the presence of shared ancestry) and partially unresolved mutational pathways.

31. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
32. Petschenka, G., Pick, C., Wagschal, V. & Dobler, S. Functional evidence for physiological mechanisms to circumvent neurotoxicity of cardenolides in an adapted and a non-adapted hawk-moth species. *Proc. R. Soc. B* **280**, 20123089 (2013).
33. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
34. Pupko, T., Pe'er, I., Shamir, R. & Graur, D. A fast algorithm for joint reconstruction of ancestral amino acid sequences. *Mol. Biol. Evol.* **17**, 890–896 (2000).
35. Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
36. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
37. Levy Karin, E., Ashkenazy, H., Wicke, S., Pupko, T. & Mayrose, I. TraitRateProp: a web server for the detection of trait-dependent evolutionary rate shifts in sequence sites. *Nucleic Acids Res.* **45**, W260–W264 (2017).
38. Poon, A. F., Lewis, F. I., Pond, S. L. & Frost, S. D. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLOS Comput. Biol.* **3**, e231 (2007).
39. Gratz, S. J. et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* **196**, 961–971 (2014).
40. Ponton, F., Chapuis, M. P., Pernice, M., Sword, G. A. & Simpson, S. J. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*. *J. Insect Physiol.* **57**, 840–850 (2011).
41. Taussky, H. H. & Shorr, E. A microcolorimetric method for the determination of inorganic phosphorus. *J. Biol. Chem.* **202**, 675–685 (1953).
42. Petschenka, G. & Dobler, S. Target-site sensitivity in a specialized herbivore towards major toxic compounds of its host plant: the $Na^+K^+$-ATPase of the oleander hawk moth (*Daphnis nerii*) is highly susceptible to cardenolides. *Chemoecology* **19**, 235–239 (2009).
43. Beikirch, H. Toxicity of ouabain on *Drosophila melanogaster*. *Experientia* **33**, 494–495 (1977).
44. Ja, W. W. et al. Prandiology of *Drosophila* and the CAFE assay. *Proc. Natl Acad. Sci. USA* **104**, 8253–8256 (2007).
45. Glass, H. W. Jr & Pan, M. L. Laboratory rearing of monarch butterflies (Lepidoptera: Danaiidae), using an artificial diet. *Ann. Entomol. Soc. Am.* **76**, 475–476 (1983).
46. Malcolm, S. B. & Zalucki, M. P. Milkweed latex and cardenolide induction may resolve the lethal plant defence paradox. *Entomol. Exp. Appl.* **80**, 193–196 (1996).
47. Šali, A., Potterton, L., Yuan, F., van Vlijmen, H. & Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318–326 (1995).
48. Laursen, M., Yatime, L., Nissen, P. & Fedosova, N. U. Crystal structure of the high-affinity $Na^+K^+$-ATPase-ouabain complex with $Mg^{2+}$ bound in the cation binding site. *Proc. Natl Acad. Sci. USA* **110**, 10958–10963 (2013).
49. Gregersen, J. L., Mattle, D., Fedosova, N. U., Nissen, P. & Reinhard, L. Isolation, crystallization and crystal structure determination of bovine kidney $Na^+$,$K^+$-ATPase. *Acta Crystallogr. F* **72**, 282–287 (2016).
50. Janson, G., Zhang, C., Prado, M. G. & Paiardini, A. PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics* **33**, 444–446 (2017).
51. The PyMOL molecular graphics system v.1.8 (Schrödinger LLC, 2015).
52. Morris, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
53. Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* **17**, 57–61 (1999).
54. Ganetzky, B. & Wu, C. F. Indirect suppression involving behavioral mutants with altered nerve excitability in *Drosophila melanogaster*. *Genetics* **100**, 597–614 (1982).

**Acknowledgements** We thank V. Wagschal, who helped with the construction and testing of in vitro cell lines; E. Toprak, who provided MATLAB code; T. O'Connor, who aided in the sequestration analyses; the Essig Museum of Entomology for photographs of milkweed butterfly specimens; M. Fa and K. O'Connor-Giles for advice on the development of the fly lines; and E. LaPlante for assistance with feeding assays. D. Bachtrog, K. Mooney, P. Moorjani, M. Nachman, R. Nielsen, P. Sudmant, R. Tarvin and B. Walsh provided feedback that improved the manuscript. Access to the HPC resources of Aix-Marseille Université was supported by the Equip@Meso (ANR-10-EQPX-29-01) project of the Investissements d'Avenir supervised by the Agence Nationale de la Recherche. This project was supported by grants from the Gordon and Betty Moore Foundation (Life Sciences Research Foundation Postdoctoral Fellowship Grant GBMF2550.06 to S.C.G.), the German Research Foundation (DFG, grant Do527/5-1 to S.D.), the Agence National de la Recherche (grant no. BioHSFS ANR-15-CE11-0007 to F.S. and F.R.), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 772257 to F.S. and F.R.), the National Geographic Society (grant 9097-12 to N.K.W.), the National Science Foundation (grant DEB-1256758 to N.K.W. and IOS-1907491 to A.A.A.), the John Templeton Foundation (grant ID 41855 to A.A.A., S.D., and N.K.W.), and the National Institute of General Medical Sciences of the National Institutes of Health (award no. R35GM119816 to N.K.W.).

**Author contributions** M.K. co-designed and implemented the overall strategy for the creation of the knock-in fly lines, designed and implemented the bioassays, the RT–qPCR experiments and the RMO analysis, performed statistical analyses and co-wrote the manuscript. S.C.G. designed and implemented the overall strategy for the creation of the knock-in fly lines, prepared the sequence data and metadata for the phylogenetic analyses, co-designed all other experiments, and co-wrote the manuscript. F.S. performed the structural modelling and docking site analyses. J.N.P. performed the phylogenetic, ancestral state and co-evolutionary analyses. K.I.V. conducted crosses, genotyping, and feeding experiments, and co-designed the qPCR experiments. J.M.A. and S.L.B. conducted crosses and genotyping, and feeding and sequestration experiments. A.P.H. performed the in vitro physiological experiments and sequestration analyses. T.M. conducted feeding experiments M.A. performed the RMO analysis with M.K., and conducted genotyping and feeding experiments. G.G. completed the RMO and ouabain dietary survival analyses. F.R. supervised the structural modelling and docking site analyses. S.D. oversaw and interpreted in vitro cell line analyses, helped to design the overall project and co-wrote the manuscript. A.A.A. helped to design the overall project, oversaw the in vitro physiological and sequestration experiments, and co-wrote the manuscript. N.K.W. led the overall collaboration, the project design and its integration, creation of fly lines and statistical analyses, and co-wrote the manuscript.

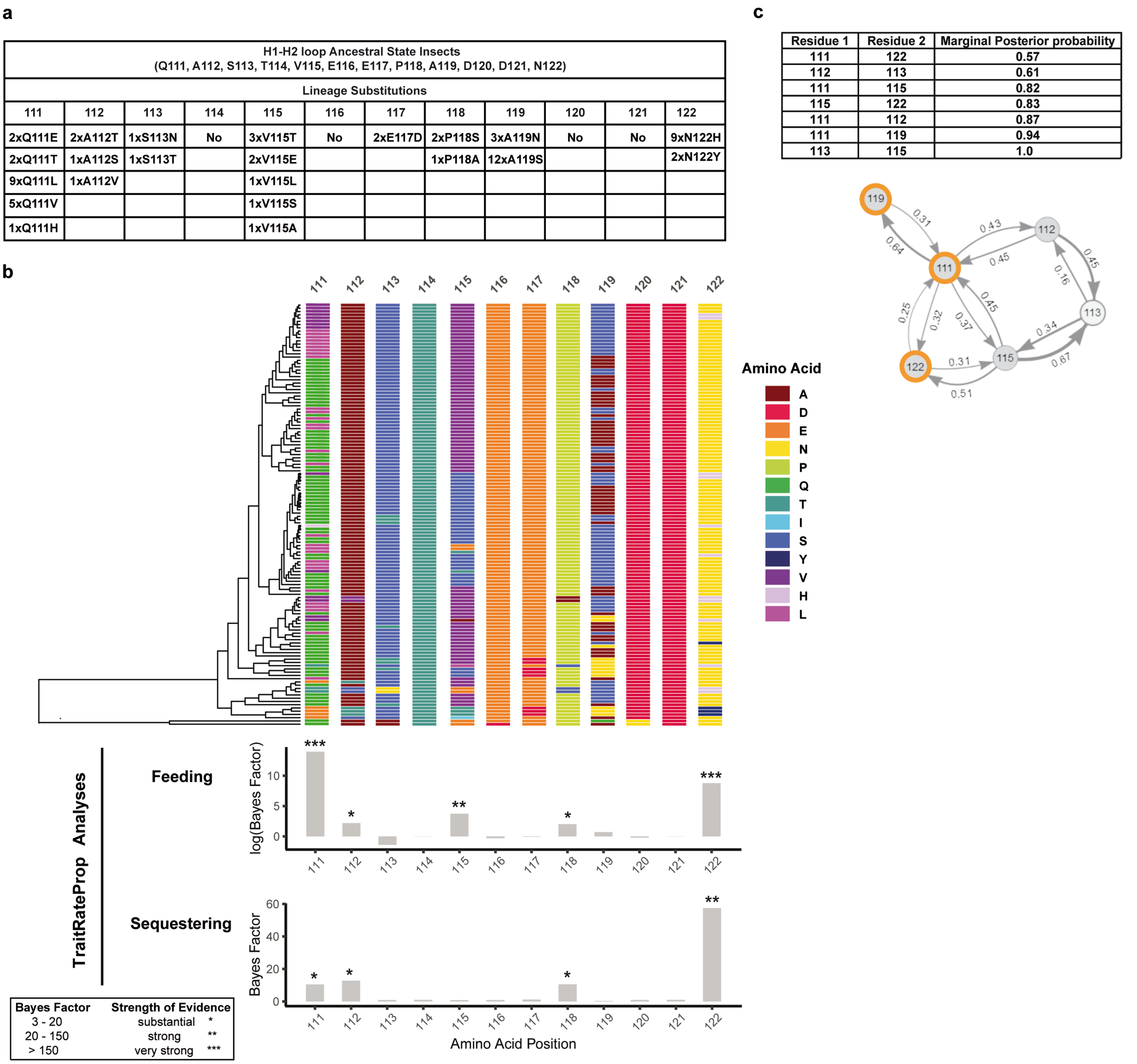
**Competing interests** The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1610-8.

**Correspondence and requests for materials** should be addressed to N.K.W.

**Peer review information** *Nature* thanks Joseph W. Thornton and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



**a**

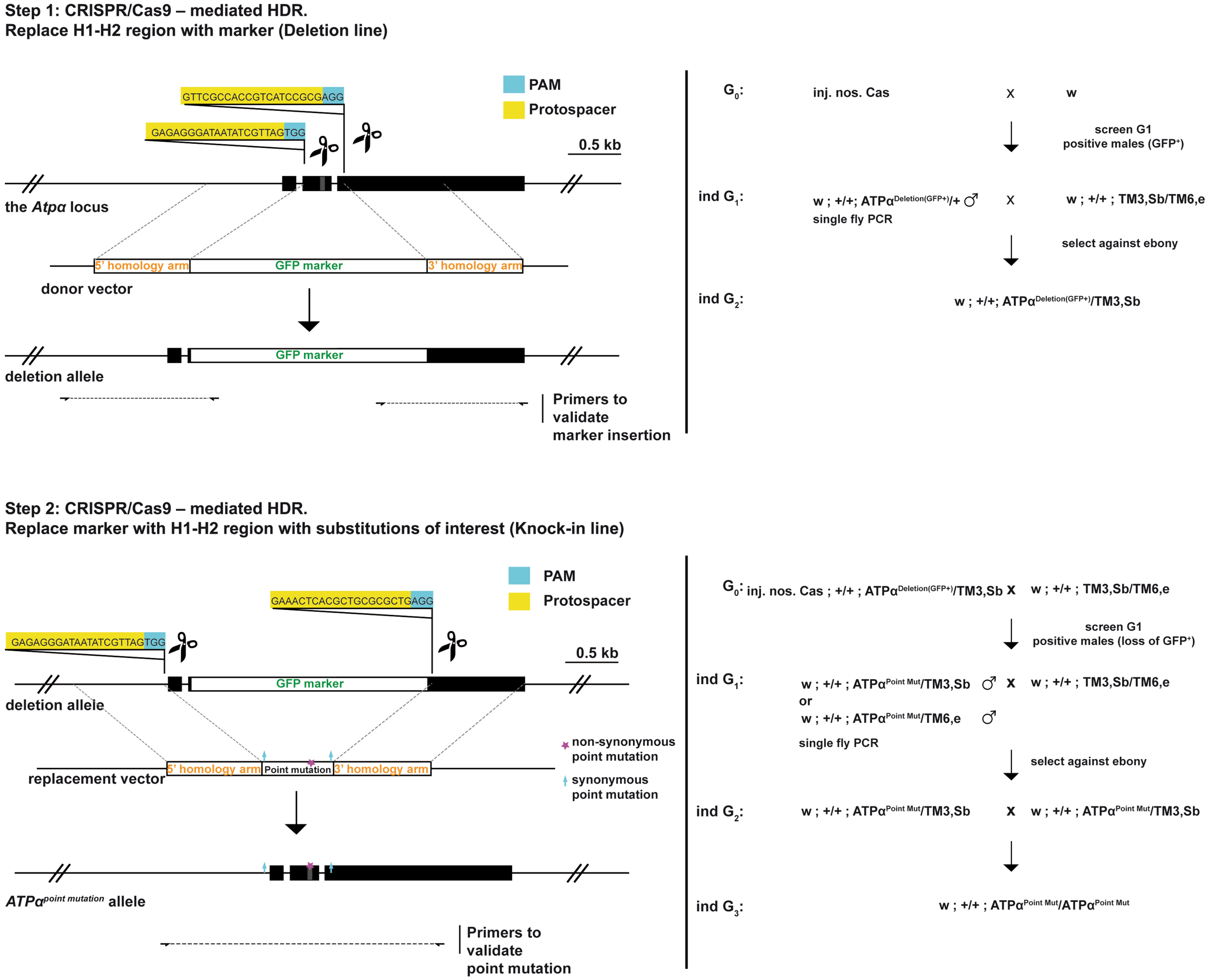
| H1-H2 loop Ancestral State Insects<br>(Q111, A112, S113, T114, V115, E116, E117, P118, A119, D120, D121, N122) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Lineage Substitutions** | | | | | | | | | | | |
| 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 |
| 2xQ111E | 2xA112T | 1xS113N | No | 3xV115T | No | 2xE117D | 2xP118S | 3xA119N | No | No | 9xN122H |
| 2xQ111T | 1xA112S | 1xS113T | | 2xV115E | | | 1xP118A | 12xA119S | | | 2xN122Y |
| 9xQ111L | 1xA112V | | | 1xV115L | | | | | | | |
| 5xQ111V | | | | 1xV115S | | | | | | | |
| 1xQ111H | | | | 1xV115A | | | | | | | |

**c**

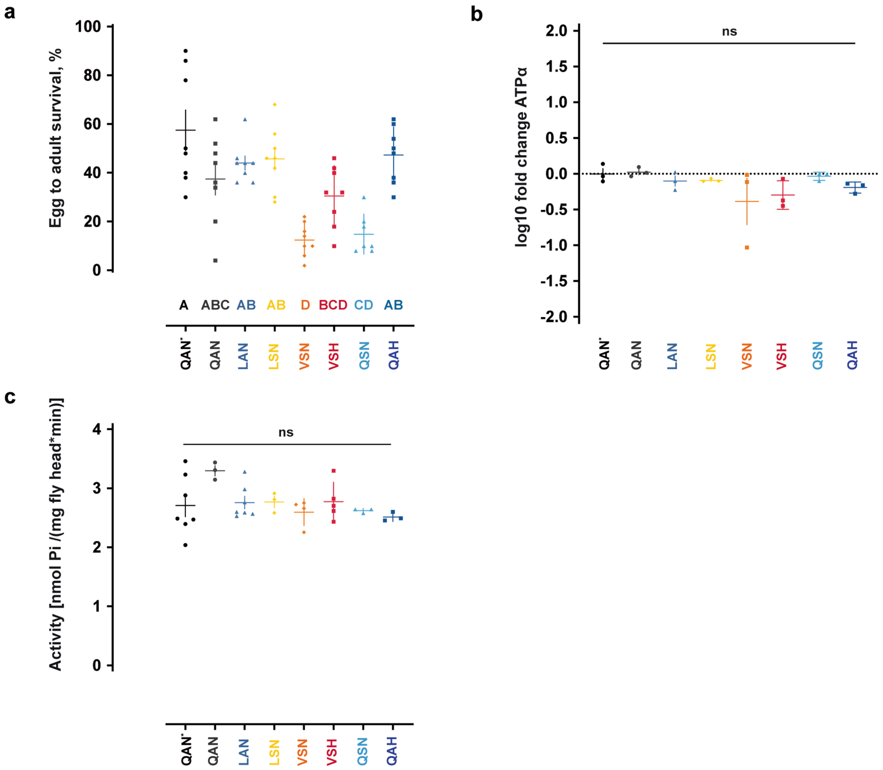
| Residue 1 | Residue 2 | Marginal Posterior probability |
|---|---|---|
| 111 | 122 | 0.57 |
| 112 | 113 | 0.61 |
| 111 | 115 | 0.82 |
| 115 | 122 | 0.83 |
| 111 | 112 | 0.87 |
| 111 | 119 | 0.94 |
| 113 | 115 | 1.0 |

| Bayes Factor | Strength of Evidence |
|---|---|
| 3 - 20 | substantial * |
| 20 - 150 | strong ** |
| > 150 | very strong *** |

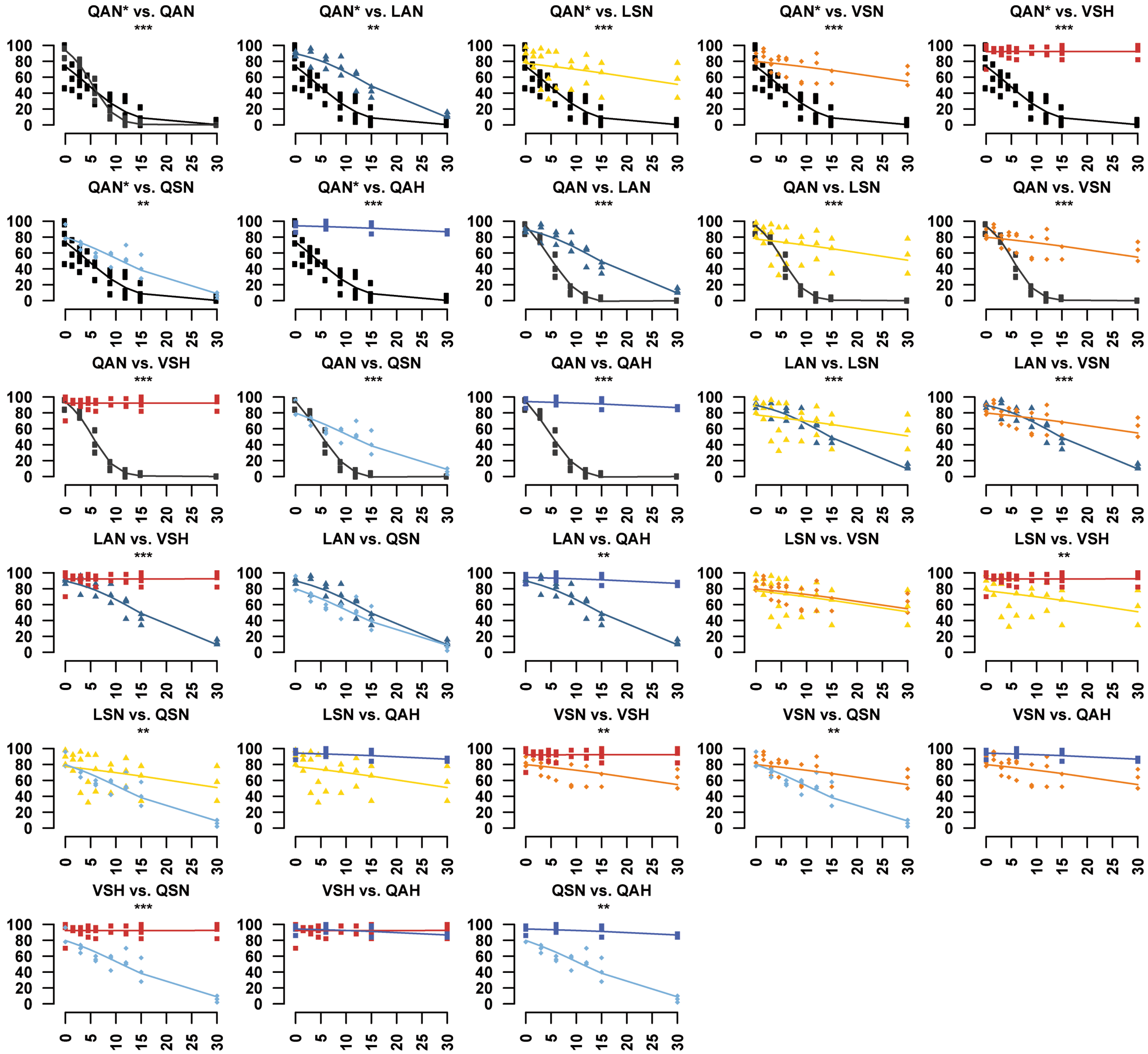
**Extended Data Fig. 1 | Substitutions at ATP$\alpha$ amino acid residues 111, 119 and 122 are directly or indirectly associated with insect specialization on plants that produce cardiac glycosides. a**, The number of occurrences of each substitution across the 21 lineages in which specialization evolved independently. **b**, TraitRateProp analysis of the H1–H2 loop of ATP$\alpha$ across insects shows amino acid residues that are strongly associated with feeding on cardiac glycoside-producing plants and toxin sequestration. Bayes factor values in the top histogram indicate per-site associations between feeding and sequence rate evolution, values in the bottom histogram indicate per-site associations between sequestration and sequence rate evolution. Values over 10 were considered different (asterisks). For information on the species included in the analysis please see Supplementary Text. Colours in the multi-sequence alignment represent individual amino acids. **c**, BGM shows the correlated evolution of amino acid sites within the H1–H2 loop of ATP$\alpha$. The table shows the marginal posterior probabilities (PP) between amino acid interactions, where the PP exceeds a default cut-off of 0.5. The residue interactions are depicted graphically, with amino acid sites represented by the nodes and the PP associated with a given epistatic or co-evolutionary interaction indicated by the values at the arrows. Nodes circled in orange indicate amino acid sites that are the focus of experiments in this study. Sites 111 and 122 are very strongly associated with feeding and sequestering, and site 119 co-evolves with site 111.



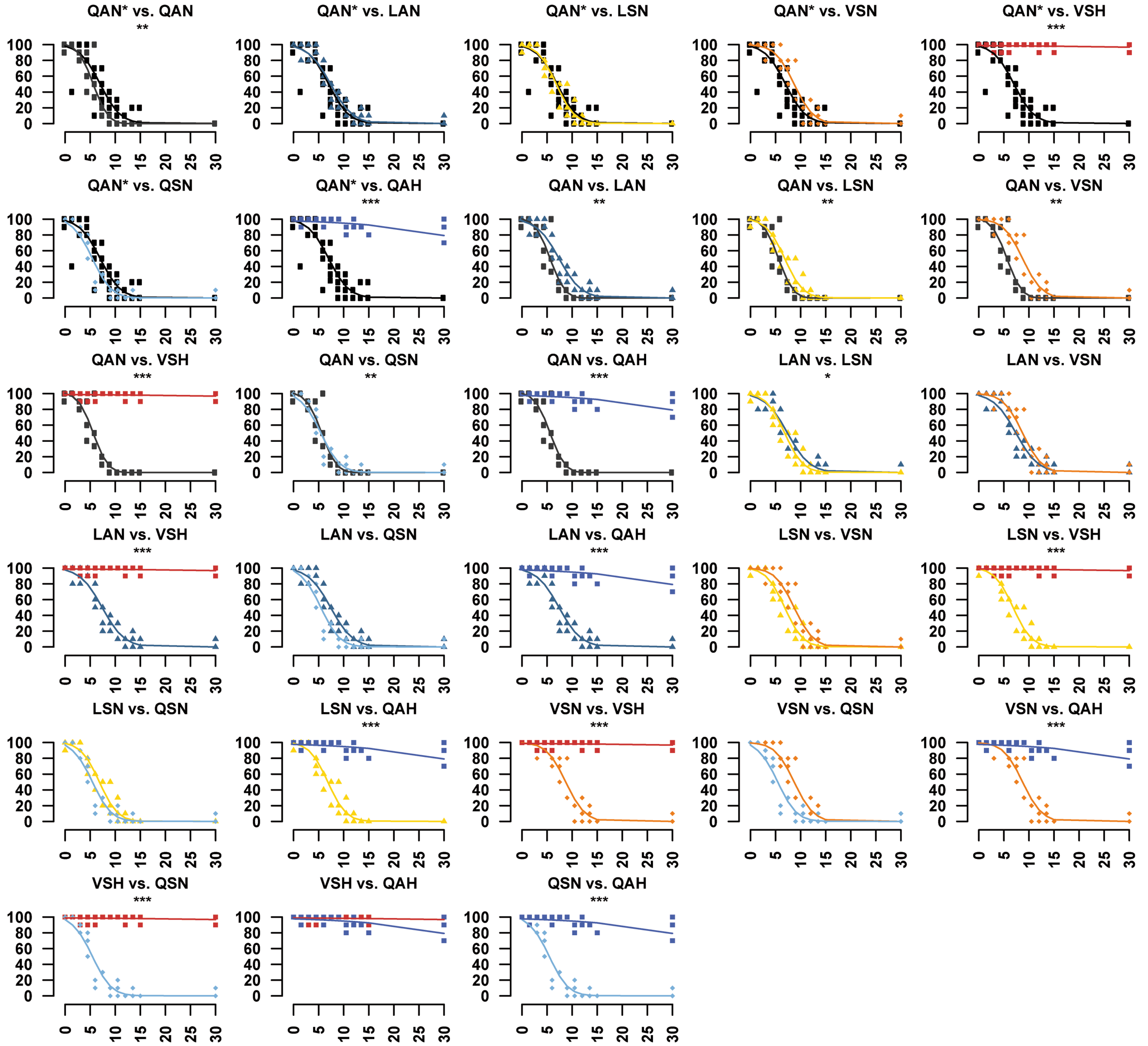
**Extended Data Fig. 2 | A two-step genome editing approach using CRISPR–Cas9 and HDR to generate knock-in *Atpα* lines of *Drosophila melanogaster*.** In the first step, the region encoding the H1–H2 extracellular domain was replaced with a 3 × P3::GFP marker through CRISPR–Cas9-mediated HDR. This generated a common parent line with the deletion allele *Atpα*$^{\text{Deletion (GFP+)}}$. In the second step, the 3 × P3::GFP marker was replaced with each of the synonymous and non-synonymous point mutation alleles through an additional round of CRISPR–Cas9-mediated HDR to generate the knock-in lines. The crossing schemes to establish the deletion line and the knock-in lines following the first and second rounds of HDR, respectively, are also shown. See also Methods and Supplementary Tables 4–7 for further details on the genome engineering strategy and crosses behind the establishment of the knock-in lines.



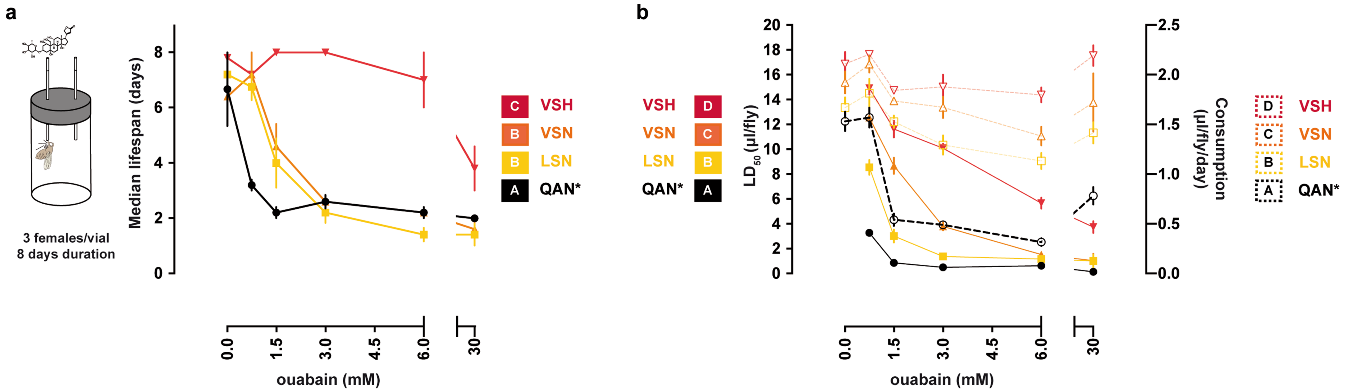
**Extended Data Fig. 3 | Point mutations have some effect on adult emergence, but do not lead to major changes in baseline *Atpα* expression or Na$^+$/K$^+$-ATPase activity. a**, Percentages of emerging adults of the knock-in and control lines on standard *Drosophila* medium ($n = 7$–8 vials, 100 eggs per vial, mean ± s.e.m.). Survival of the knock-in lines and control lines QAN (engineered control) and QAN* ($w^{1118}$ wild type) was compared using one-way ANOVA ($P < 0.001$). Survival differed between QAN* and some of the knock-in lines, but not between the engineered control line QAN and any of the knock-in lines except VSN (post hoc Tukey's tests (letters)). **b**, *Atpα* expression was not different among the engineered *Drosophila* knock-in lines or $w^{1118}$ wild-type flies (QAN*). *Atpα* transcript level differences were assayed by qPCR. Expression was assayed in three biological replicates (symbols represent the mean ± s.e.m.), with two technical replicates per biological replicate (averaged for each biological replicate), of five- to six-day-old females as fold change standardized against *rpl32* expression in QAN* flies. The expression fold change between genotypes was compared using one-way ANOVA ($P = 0.3197$). **c**, None of the sequential *Atpα* genotypes found along the monarch lineage affected base-line levels of pump activity in a sodium pump enzymatic assay using extracts of fly heads (one-way ANOVA, $P = 0.1377$; symbols represent the mean ± s.e.m. of 3–7 biological replicates). Further information on experimental design and statistical test results is in the Source Data. ns, not significant.



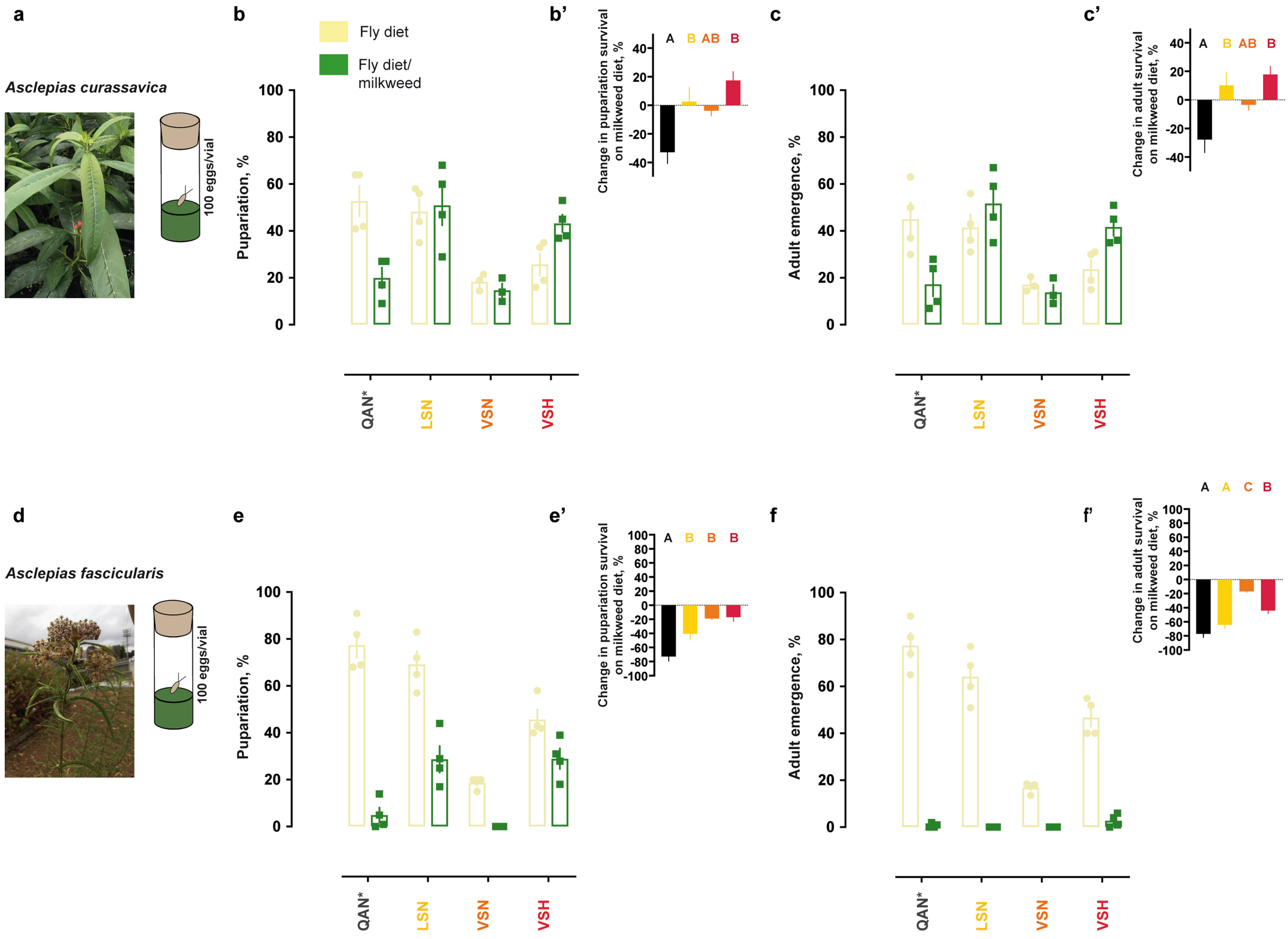
**Extended Data Fig. 4 | *Drosophila* flies with edited genomes show increased larval–adult survival when fed cardiac glycosides.** These panels accompany Fig. 2c. Larval–adult survival when reared on diets with a range of ouabain concentrations was different between monarch lineage knock-in lines relative to control lines (QAN, engineered control; QAN*, $w^{1118}$ wild type). Symbols represent the mean ± s.e.m. of 3–6 biological replicates (50 larvae per replicate). Curves were fit through a univariate logistic regression (effect of ouabain concentration on survival), and the difference in survivorship trajectories between each pair of fly lines (genotypes) was evaluated by performing an LRT to assess the significance of the inclusion of an interaction term between genotype and ouabain concentration in the logistic regression for a pair of lines (**$P < 0.01$, ***$P < 0.001$). Further information on experiment design and statistical test results is in the Source Data.



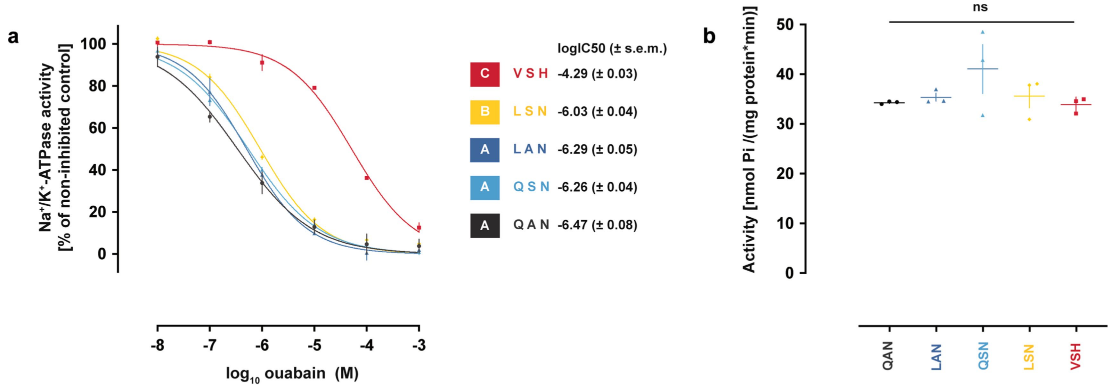
**Extended Data Fig. 5 | *Drosophila* flies with edited genomes show increased adult survival when fed cardiac glycosides.** These panels accompany Fig. 2d. Adult survival when reared on diets with a range of ouabain concentrations was different between monarch lineage knock-in lines and control lines (QAN, engineered control; QAN*, $w^{1118}$ wild type). Symbols represent the mean ± s.e.m. of 3–6 biological replicates. Curves were fit through a univariate logistic regression (the effect of ouabain concentration on survival), and a difference in survivorship trajectories between each pair of fly lines (genotypes) was evaluated by performing an LRT to assess the significance of the inclusion of an interaction term between genotype and ouabain concentration in the logistic regression for a pair of lines ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$). Further information on experimental design and statistical test results is in the Source Data.



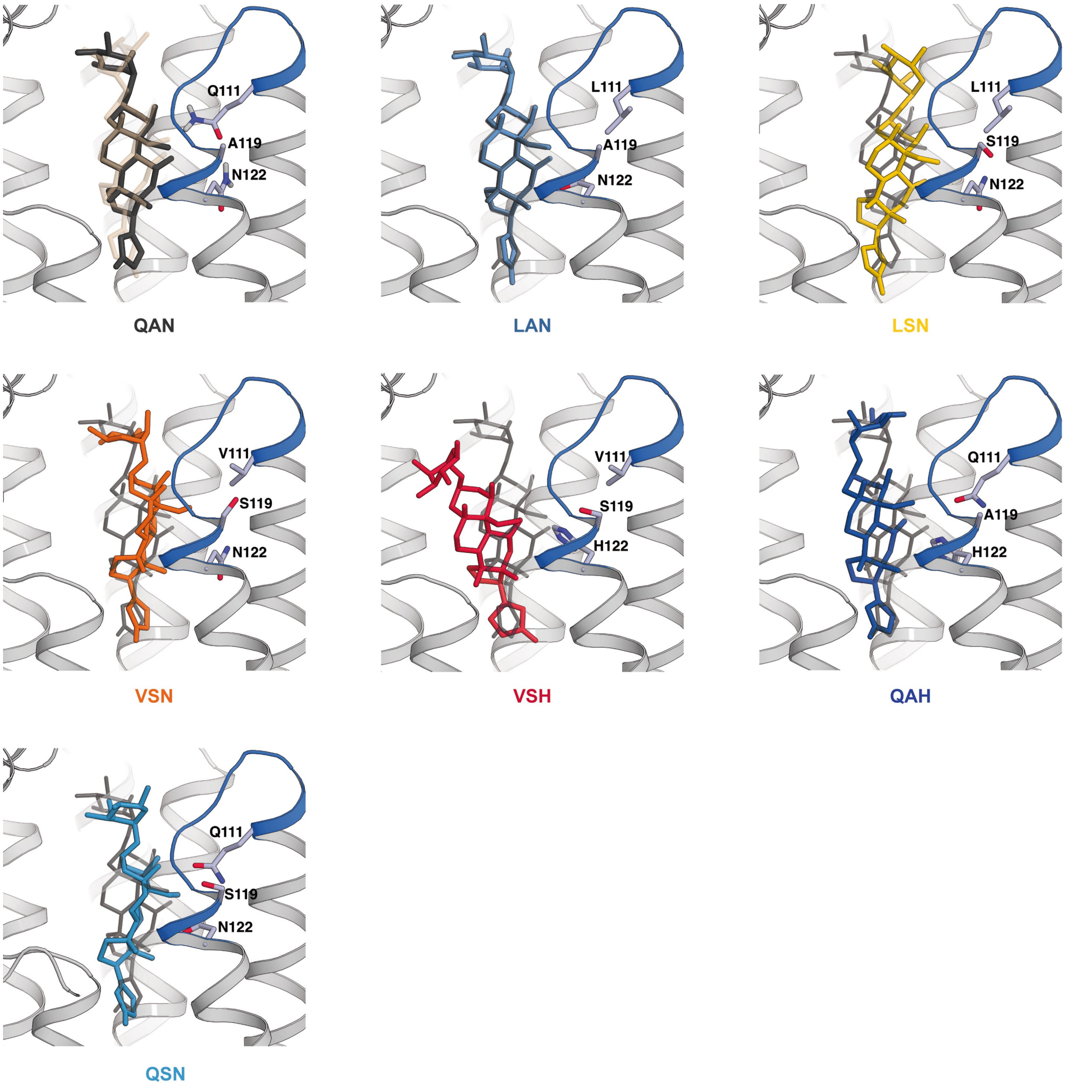
**Extended Data Fig. 6 | The mutational path of ATPα in the monarch butterfly lineage increases dietary tolerance to ouabain in vivo in engineered *Drosophila* without affecting feeding rate. a**, Estimation of mean lifespan (days) in adult females (four to seven days old at the start of the experiment) of the knock-in and control lines in CAFE assays across a range of ouabain concentrations. Each data point represents the mean ± s.e.m. of five biological replicates. Both ouabain concentration and genotype affect the survival time (two-way ANOVA ($P < 0.0001$) with post hoc Tukey's tests (letters indicate pairwise differences between genotypes)). **b**, Estimation of $LD_{50}$ (µl per fly; solid lines**)** and feeding rates (µl per fly per day; dashed lines) in the same individuals as in **a**. Each data point represents the mean ± s.e.m. of five biological replicates. Both ouabain concentration and genotype affect $LD_{50}$ (two-way ANOVA ($P < 0.0001$) with post hoc Tukey's tests (letters indicate pairwise differences between genotypes)). Further information on experimental design and statistical test results is in the Source Data.



**Extended Data Fig. 7 | Survival of knock-in lines on fly diets supplemented with dried, pulverized leaves of two milkweed species that host monarch butterflies in nature (*A. curassavica* and *A. fascicularis*).** **a**, Photograph of *A. curassavica* plant used in this study. **b**, **c**, Percentages of pupariating larvae and emerging adults of the knock-in and control (wild-type $w^{1118}$: QAN*) lines on fly diet with and without *A. curassavica* leaf material ($n = 3$–4, mean ± s.e.m.). **b′**, **c′**, Differences in pupariation and emergence percentages on a fly diet with milkweed relative to percentages on a control diet ($n = 3$–4, mean ± s.e.m.). Mean differences between percentages in **b′** and **c′** were tested with one-way ANOVA ($P < 0.01$) followed by post hoc Tukey's tests (letters). These panels accompany Fig. 2e. **d**, Photograph of *A. fascicularis* plant used in this study. **e**, **f**, Percentages of pupariating larvae and emerging adults of the knock-in and control lines on fly diet with and without *A. fascicularis* leaf material ($n = 4$, mean ± s.e.m.). **e′**, **f′**, Differences in pupariation and emergence percentages on a fly diet with milkweed relative to percentages on a control diet ($n = 4$, mean ± s.e.m.). Mean differences between percentages in **e′** and **f′** were tested with one-way ANOVA ($P < 0.001$) followed by post hoc Tukey's tests (letters). Experiments were performed once, and adding leaf material of either of the two milkweed species to the fly diet had largely consistent effects on survival of the monarch lineage knock-in and control fly lines.



**Extended Data Fig. 8 | The *Atpα* genotypes found along the monarch lineage sequentially increase TSI to ouabain without affecting baseline levels of sodium pump activity. a**, In vitro ouabain sensitivity of engineered *Drosophila* $Na^+/K^+$-ATPases transiently expressed in Sf9 cell lines. Each data point represents the mean ± s.e.m. of three biological replicates. $\log_{10}[IC_{50}]$ for each type of $Na^+/K^+$-ATPase was estimated after four-parameter logistic curve fitting, and statistical differences between $\log_{10}[IC_{50}]$ values were tested with one-way ANOVA ($P < 0.0001$) followed by post hoc Tukey's tests (letters). **b**, None of the sequential *Atpα* genotypes found along the monarch lineage affected baseline levels of pump activity in the enzymatic assay with extracts of transiently transfected Sf9 cells (one-way ANOVA, $P = 0.3197$). Each data point represents the mean ± s.e.m. of three biological replicates.



**Extended Data Fig. 9 | Molecular docking simulations show stepwise reductions in ouabain binding to Na$^+$/K$^+$-ATPases with monarch lineage substitutions in ATP$\alpha$.** The ouabain binding pocket structure obtained from molecular docking simulations for each Na$^+$/K$^+$-ATPase with mutated ATP$\alpha$. The mutated residues are shown in sticks and are labelled. The H1–H2 loop of ATP$\alpha$ is shown in blue. The extracellular region of the $\alpha$-subunit is removed for simplicity. For the wild-type (QAN) ATPase, ouabain is shown in its co-crystal structure coordinates (white, transparent) together with its best-docked position. For all other ATPases only the best-docked positions (closest to the co-crystal structure) are shown together with ouabain's docking position for the wild-type ATPase (dark grey). The triple-mutated VSH ATPase has two distinct docking scores: one is similar to the docking energy for the wild-type ATPase and the other has the lowest binding energy compared to all other mutated ATPases. The potential existence of both states might be related to a trend of reduced bang sensitivity for this genotype compared to some of the single-mutant genotypes. A119 is not directly part of the ouabain binding pocket, and therefore, A119S alone does not change ouabain binding. Although the consequences of A119S are relatively subtle, the mutation may disrupt the local hydrogen bonding network and cause structural or dynamic changes in the loop or in its vicinity.

nature research

Corresponding author(s): Noah K. Whiteman

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on statistics for biologists may be useful.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | The StepOne™ Real-Time PCR System (ThermoFisher Scientific) was used for qPCR. The Agilent 1100 series HPLC unit (Agilent, Santa Clara CA) was used to detect ouabain in the sequestration assay. |
| Data analysis | Tree construction was performed with IQ-TREE (Trifinopoulos et al. 2016). Analysis of residue associations with feeding was performed with TraitRateProp (Karin et al. 2017), and the makeChronosCalib function of the ape package in R (Paradis et al., 2004). Coevolving sites analysis was performed with Spidermonkey/BGM in HyPhy (Poon et al. 2007). Ancestral sequence reconstruction was performed in HyPhy (Pupko et al. 2000; Pond et al. 2005). Feeding and sequestering ancestral state reconstruction was performed with the ace function of the ape package in R (Paradis et al., 2004). The tree and character states were visualized using Interactive Tree Of Life (iTOL) v3 (Letunic et al., 2016), and can be found here: https://itol.embl.de/tree/1361521431149621561136381. Origin-Pro 9.1G was used to determine the Hill plots and coefficients. Statistical analysis was performed using Prism 8 (Graphpad) or R (including some analyses based on code developed originally for MATLAB). Protein homology modeling and in silico mutagenesis was performed via Modeller (Sali and Blundell, 1993), PyMod2 (Janson et al., 2016), ClustalW and PyMOL (Schrödinger, 2015). Molecular docking simulations were performed with Autodock4 (Morris et al. 1998; Huey et a. 2007). Statistical analysis to assess the reproducibility of the mutational order was performed with a modified custom algorithm in MATLAB (Toprak et al., 2011) kindly provided ny Prof. Erdal Toprak. Please refer to Methods for details. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data for all bioassays are provided in the associated Source data excel file of each Figure. Correspondence and requests for other materials should be addressed to N.K.W. (whiteman@berkeley.edu)

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | We performed preliminary experiments to assess variance and estimate sample sizes before conducting all bioassays, following sample sizes published for studies conducting identical or similar experiments, most of which were papers that included present co-authors. For feeding experiments, we relied on previously published studies using feeding experiments with ouabain and Drosophila. Although our sample sizes for feeding experiments on fly media were 10-200 flies per vial for the various feeding experiments, a sample size of N = 3 per vial with 3 biological replicate vials was used for the CAFE assays, and this was chosen because similar assays were reported in a paper published by us in which we used feeding experiments with ouabain and Drosophila mutants in order to assess survival (Groen et al., 2017, doi: 10.1016/j.ibmb.2016.12.008). For other feeding experiments sample sizes ranged from N = 10 flies per vial to 200 eggs per vial (an over-abundance was used because of variance in survival across knock-in lines at eclosion that we report in the manuscript), and were thus conservative. For enzyme assays using insect heads, we followed previously published sample sizes for the number of heads per biological replicate (Petschenka et al., 2013, doi: 10.1111/evo.12152), and similarly for enzyme assays using Sf9 cell lines (Dalla and Dobler, 2016 ,doi: 10.1111/evo.13077). For sequestration analyses, we settled on N = 2-4 biological replicates given similar experiments in citations above, and each biological replicate contained pools of 50-100 individuals based on preliminary studies with standards. |
| Data exclusions | In the CAFE assay the vials in which all flies died early in combination with a lack of feeding were excluded from all analyses. These deaths were attributable to flies being unable to access the food. Further vials were excluded from LD50 and feeding rate measurements if there was a lack of feeding for one of the days (again the result of flies being unable to access the food) or if there were missing feeding data. These were pre-established exclusion criteria. |
| Replication | We performed all bioassays as well as enzyme assays in replicates. Results of all assays were reproducible. Please refer to Methods for numbers of replicates. |
| Randomization | We used random.org/sequences to randomize placement of tubes in racks for all bioassays, except for the feeding experiments with milkweed and the sequestration assays. For the milkweed feeding experiments, the initial experiment was done as a preliminary study without randomization and we did not anticipate it working because of the myriad compounds in the plants, which we thought would kill even the monarch genotype flies. The monarch flies, to our surprise, survived relatively well on media containing milkweed, but because milkweed was no longer available (we collected it on UC-Berkeley campus), except enough for one repeat of the experiment, we followed the first design (no randomization) to keep our design consistent. For the sequestration assays, puparia were transfered out of the vials to prevent freshly emerging flies from feeding on the media containing ouabain, after which flies were collected flies every two hours. For these reasons we did not randomize. |
| Blinding | We were blind to the allocation and outcome assessment during all bioassays, except for the feeding experiments with milkweed and the sequestration assays. For the milkweed experiments, the media is green and thus reveals to the experimenter which vial is in which treatment group. For the sequestration assays, the personnel who conducted the rearing were at a different university (Cornell University) than those who conducted the HPLC. In addition, we had to know which flies emerged from which vials to group them and freeze them. Given these procedures, we were not blind to the experimental treatment for these two experiments. |

# Reporting for specific materials, systems and methods

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | Unique biological materials |
| ☒ | ☐ | Antibodies |
| ☐ | ☒ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology |
| ☐ | ☒ | Animals and other organisms |
| ☒ | ☐ | Human research participants |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | ChIP-seq |
| ☒ | ☐ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

# Unique biological materials

Policy information about availability of materials

| | |
|---|---|
| Obtaining unique materials | All unique materials are available upon request. The knock-in fly lines will be available from the Bloomington Drosophila Stock Center or directly from N.K.W. (email to whiteman@berkeley.edu). |

# Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | Sf9 cells in Grace's were originally purchased from Gibco™ , now owned by ThermoFisher, under catalog number B82501. |
| Authentication | From ThermoFisher's website (https://www.thermofisher.com/order/catalog/product/B82501?SID=srch-srp-B82501): "Each lot of Gibco® Sf9 cells is tested for cell growth and viability post-recovery from cryopreservation...and has been characterized by isozyme and karyotype analysis." In addition, the cells have been maintained at the University of Hamburg (Susanne Dobler) for the last 12 years and checked regularly for contamination by bacteria and mycoplasma. |
| Mycoplasma contamination | From ThermoFisher's website (https://www.thermofisher.com/order/catalog/product/B82501?SID=srch-srp-B82501): "In addition, the Master Seed Bank has been tested for contamination of bacteria, yeast, mycoplasma and virus..." |
| Commonly misidentified lines<br>(See ICLAC register) | Cells are not listed in the database. |

# Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | Only invertebrate animals were used in our experiments. We used eggs, larvae (1st-3rd instar), puparia and adults (females, aged 4-7 days) of the w1118 strain and the 8 knock-in lines we created in Drosophila melanogaster (Diptera, Drosophilidae) for all experiments. Please refer to the methods section for a description of the transgenics and genetics methods and validation. The Berkeley campus Committee on Laboratory & Environmental Biosafety has approved a Biological Use Authorization (#451) for the maintenance and disposal of the knock-in D. melanogaster lines generated by GenetiVision (Houston, TX), adhering to regulations of the National Institutes of Health in the laboratory of Prof. Noah Whiteman. |
| Wild animals | We did not use wild animals for the present study. |
| Field-collected samples | We did not use field-collected samples for the present study. |

LETTER

https://doi.org/10.1038/s41586-019-1641-1

# Population imaging of neural activity in awake behaving mice

Kiryl D. Piatkevich[1,2,11], Seth Bensussen[3,11], Hua-an Tseng[3,11], Sanaya N. Shroff[3], Violeta Gisselle Lopez-Huerta[4], Demian Park[1,2], Erica E. Jung[1,5], Or A. Shemesh[1,2], Christoph Straub[6], Howard J. Gritton[3], Michael F. Romano[3], Emma Costa[1], Bernardo L. Sabatini[6], Zhanyan Fu[4], Edward S. Boyden[1,2,7,8,9,10]* & Xue Han[3]*

**A longstanding goal in neuroscience has been to image membrane voltage across a population of individual neurons in an awake, behaving mammal. Here we describe a genetically encoded fluorescent voltage indicator, SomArchon, which exhibits millisecond response times and is compatible with optogenetic control, and which increases the sensitivity, signal-to-noise ratio, and number of neurons observable several-fold over previously published fully genetically encoded reagents[1–8]. Under conventional one-photon microscopy, SomArchon enables the routine population analysis of around 13 neurons at once, in multiple brain regions (cortex, hippocampus, and striatum) of head-fixed, awake, behaving mice. Using SomArchon, we detected both positive and negative responses of striatal neurons during movement, as previously reported by electrophysiology but not easily detected using modern calcium imaging techniques[9–11], highlighting the power of voltage imaging to reveal bidirectional modulation. We also examined how spikes relate to the subthreshold theta oscillations of individual hippocampal neurons, with SomArchon showing that the spikes of individual neurons are more phase-locked to their own subthreshold theta oscillations than to local field potential theta oscillations. Thus, SomArchon reports both spikes and subthreshold voltage dynamics in awake, behaving mice.**

Near-infrared genetically encoded voltage indicators (GEVIs) derived from rhodopsins offer high temporal fidelity, and are compatible with optogenetics[1,12,13], whereas green fluorescent GEVIs derived from the voltage-sensing domains of phosphatases or opsins are often slower and brighter[2,3,14–17]. Translating these voltage sensors into the living mammalian brain has been challenging, because of poor membrane localization, low photostability, and low signal-to-noise ratio (SNR). So far, amongst fully genetically encoded reagents, only Ace2N and paQuasAr3-s have been used to optically report voltage dynamics in a living mouse brain, reporting the activities of up to four cells in one field of view (FOV) in awake mice[4,17]. Recently, we developed a robotic directed-evolution approach and created the improved GEVI Archon1[13]. To further improve SNR in the dense, living mammalian brain, we conducted a screen for peptides to localize Archon1 to the soma[18–21], so that neuropil contamination could be reduced (Extended Data Fig. 1; see Supplementary Table 1 for the sequences of the motifs). The molecule Archon1–KGC–EGFP–$K_V$2.1-motif–ER2, which we call SomArchon (Fig. 1a), exhibited the highest relative change in fluorescence ($\Delta F/F$) during 100-mV voltage steps (Fig. 1g) and was well-localized to the soma (Extended Data Fig. 1h–k).

SomArchon fluorescently reported action potentials in mouse brain slices after in utero electroporation (IUE) into the cortex and hippocampus, and after adeno-associated virus (AAV)-mediated expression in the cortex, striatum, and thalamus (Extended Data Fig. 2). SomArchon was localized primarily to the membrane within 30–45 μm of the cell body in the cortex, striatum, and hippocampus (Fig. 1b, Extended Data Fig. 1h–k). The sensitivity of SomArchon was about twofold greater (Fig. 1c, d) than our previously published values for Archon1[13], and it had comparable kinetics (Fig. 1e) and SNR (defined as the maximum fluorescence change observed during an action potential divided by the standard deviation of the baseline) (Fig. 1f). SomArchon linearly reported voltage (Fig. 1g), and did not alter membrane properties or resting potential in mouse brain slices, induce gliosis, or mediate light-induced phototoxicity (Extended Data Figs. 3, 4). It has previously been demonstrated that Archon1 exhibits essentially no crosstalk under blue light illumination as used commonly for optogenetic neural activation[13]. We used a bicistronic expression system (Fig. 1h) to co-express SomArchon and the high-performance channelrhodopsin CoChR[22] in the same cell, and demonstrated that brief blue light pulses could reliably evoke action potentials that were visible in SomArchon fluorescence (Fig. 1i, j).

We performed a side-by-side comparison of SomArchon with soma-localized versions of several next-generation voltage sensors—specifically QuasAr3-s[4], paQuasAr3-s[4], ASAP3[5], and $\text{Voltron}_{525}$[6]—in mouse cortical brain slices under identical expression conditions, focusing on layer 2/3 neurons (Supplementary Table 3). The spectrally similar sensors QuasAr3-s, paQuasAr3-s, and SomArchon were recorded under identical imaging conditions (1.5 W $\text{mm}^{-2}$) during CoChR-mediated action potentials. ASAP3 and Voltron were recorded with the filter sets described[5,6] under similar excitation intensities as used in the initial description of Voltron[6] (25–29 mW $\text{mm}^{-2}$), during action potentials evoked upon application of 4-aminopyridine. Under these conditions, SomArchon exhibited the highest $\Delta F/F$ and SNR per action potential (Extended Data Fig. 5); in addition, SomArchon exhibited values higher than those previously reported for Ace2N-mNeon[7], ASAP1[2], MacQ-mCitrine[3], and QuasAr2[8] (Supplementary Table 3). In addition, SomArchon showed higher photostability than soma-localized versions of ASAP3 and $\text{Voltron}_{525}$ under comparable imaging conditions in cultured neurons (Extended Data Fig. 5e).

We virally expressed SomArchon in vivo in the mouse motor cortex, visual cortex, striatum, and hippocampus, and imaged neural activity while mice were awake with their heads fixed under a conventional one-photon microscope (Fig. 2a) using a scientific complementary metal-oxide semiconductor (sCMOS) camera and laser excitation light at 637 nm, at a power of around 1.6 W $\text{mm}^{-2}$ (75 mW; 20× objective lens), 4 W $\text{mm}^{-2}$ (75 mW; 40× objective lens), or 1.6 W $\text{mm}^{-2}$ (95 mW; 16× objective lens). Cells expressing SomArchon could be resolved at depths of about 50–150 μm below the imaging surface (Fig. 2b). We could detect individual spikes in single cells in all four brain regions (Fig. 2c, e, g, i, Supplementary Video 1). The SNR per action potential ranged from about 7 to about 16 across the brain regions examined (Fig. 2d, f, h, j). To our knowledge, no other paper has reported SNR

---

[1]Media Lab, MIT, Cambridge, MA, USA. [2]MIT McGovern Institute for Brain Research, MIT, Cambridge, MA, USA. [3]Department of Biomedical Engineering, Boston University, Boston, MA, USA. [4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Mechanical and Industrial Engineering, University of Illinois, Chicago, IL, USA. [6]Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA, USA. [7]Department of Biological Engineering, MIT, Cambridge, MA, USA. [8]MIT Center for Neurobiological Engineering, MIT, Cambridge, MA, USA. [9]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA. [10]Koch Institute, MIT, Cambridge, MA, USA. [11]These authors contributed equally: Kiryl D. Piatkevich, Seth Bensussen, Hua-an Tseng. *e-mail: esb@media.mit.edu; xuehan@bu.edu

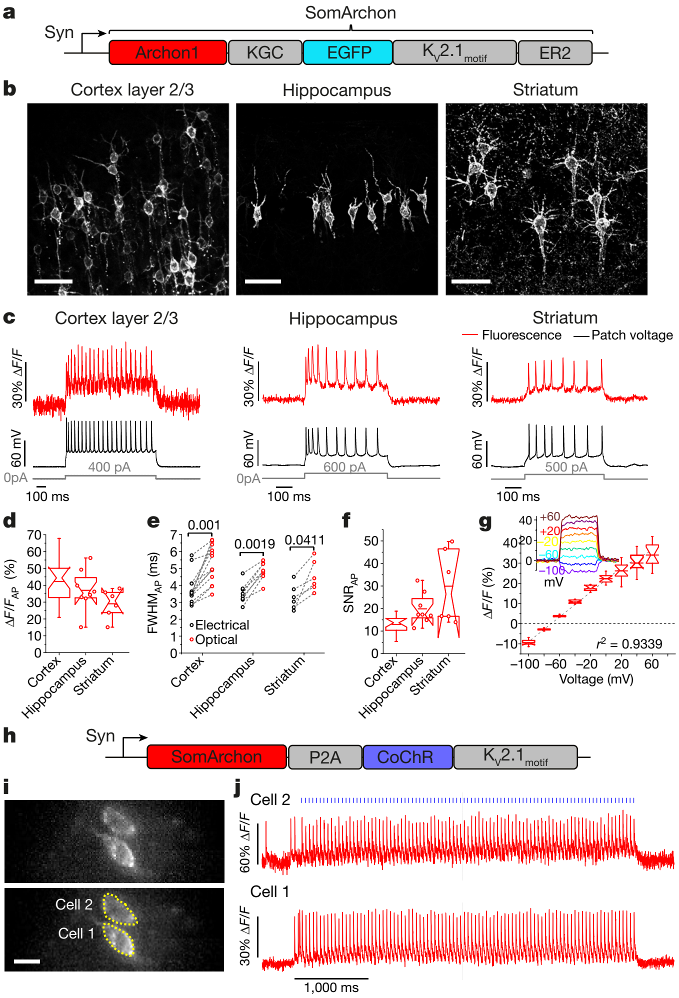


**Fig. 1 | SomArchon enables high-fidelity voltage imaging in brain slices. a**, Diagram of the SomArchon construct. **b**, Confocal images of SomArchon-expressing neurons in cortex layer 2/3 (left), hippocampus (middle), and striatum (right). Excitation wavelength ($\lambda_{ex}$) = 488 nm laser, emission wavelength ($\lambda_{em}$) = 525/50 nm (representative images selected from 8, 10 and 6 slices, respectively, from 2 mice each). Scale bars, 50 μm. **c**, Single-trial SomArchon fluorescence (red), and concurrent membrane voltage recorded via whole-cell patch-clamp (black), during action potentials evoked by current injection (grey); $\lambda_{ex}$ = 637 nm laser at 0.8, 1.5, and 1.5 W mm$^{-2}$ for cortex, hippocampus, and striatum, respectively. **d**, $\Delta F/F$ per action potential across recordings exemplified in **c** (representative traces selected from $n = 18$, 8, and 6 neurons from 5, 2, and 2 mice, respectively). Box plots: 25th and 75th percentiles with notch, median; whiskers, 1.5× interquartile range from the 25th and 75th percentiles; middle horizontal line, mean; individual data points, open circles when $n < 9$). **e**, Electrical and optical action potential waveform full-width-at-half-maximum (FWHM$_{AP}$; dashed lines connect the same neurons) across recordings exemplified in **c** ($P$ values above brackets are from two-sided Wilcoxon rank-sum test; Supplementary Table 2). **f**, SNR per action potential across recordings exemplified in **c**. **g**, Population fluorescence of SomArchon in response to voltage steps in voltage-clamp mode, normalized to the fluorescence at −70 mV (inset, optical recordings for a representative neuron) recorded in cortex ($n = 12$ neurons from 2 mice). **h**, Diagram of SomArchon–P2A–CoChR–K$_V$2.1$_{motif}$. **i**, Fluorescence image of neurons in hippocampal slice expressing SomArchon–P2A–CoChR–K$_V$2.1$_{motif}$ (top) with two cells identified (bottom); $\lambda_{ex}$ = 637 nm, exposure time 1.3 ms (selected from $n = 3$ slices from 2 mice). Scale bar, 25 μm. **j**, Representative single-trial optical voltage traces from cells shown in **i** with blue light stimulation (2 ms pulse at 20 Hz). Acquisition rate, 777 Hz.

values per action potential in living mouse brain, so we cannot directly compare our molecule to others in this regard. We were able to resolve short segments of proximal dendrites next to the soma, and detected voltage fluctuation patterns that sometimes differed from those in the soma (Extended Data Fig. 6c–f). In addition, simultaneous optical control and voltage imaging was feasible using the strategy shown in Fig. 1h, in awake behaving mice (Fig. 2k–m).

Electrophysiological studies have shown that many striatal neurons increase their responses during movement[9,23], while others decrease their responses[9,10]. Although electrophysiological recordings largely discard spatial information regarding the relative locations of the neurons being observed, recent calcium imaging studies have revealed that spatially clustered striatal neurons are activated by similar aspects of movement[11]. These calcium imaging studies focused on the increases in activity during movement; decreases in activity are harder to observe with calcium imaging. We performed voltage imaging while mice ran on a spherical treadmill (Fig. 3a), identified cell bodies and spikes from cells in the striatum (Fig. 3b–d), and aligned spiking activity to movement (Fig. 3d). Some neurons exhibited firing patterns known to occur in striatal fast spiking interneurons[24] or cholinergic interneurons[25] (Extended Data Fig. 7a–d). We found that 4 of the 14 neurons imaged were positively modulated by movement speed, and that 2 were negatively modulated by movement speed (Fig. 3e, f, Extended Data Fig. 7e, f; see Supplementary Table 2 for statistics). Furthermore, adjacent neurons did not respond to movement speed in identical ways. For example, in two recordings of three neurons each, one of the three neurons was positively modulated by movement speed whereas the other two were not (Fig. 3f, Extended Data Fig. 7a, c (cells 1, 2, 3 and cells 6, 7, 8), Supplementary Table 2). Thus, SomArchon can readily detect decreases in striatal neuron spiking during movement, and can help to disambiguate activity changes amongst spatially clustered striatal neurons.

We performed wide-field voltage imaging with SomArchon in hippocampal CA1 neurons in awake, head-fixed mice, while simultaneously recording local field potentials (LFPs). In vivo patch-clamp recordings have shown that the spikes of a CA1 neuron are more strongly phase-locked to its own intracellular theta frequency (4–10 Hz) oscillations than to the theta oscillations of the across-neuron averaged LFP[26,27]. We found that 6 of the 16 neurons had spikes phase-locked to both intracellular and LFP theta oscillations (Fig. 4a), and that 9 were phase-locked only to intracellular theta oscillations and not to LFP theta oscillations (Fig. 4b). As a population, neurons exhibited stronger phase-locking to intracellular theta oscillations than to LFP theta oscillations (Fig. 4c, d). SomArchon thus supports the analysis of subthreshold intracellular oscillations, although interpretation of these measurements must take into account background fluorescence, which—in densely labelled tissue—may result in crosstalk that affects correlations.

We evaluated the photostability of SomArchon in vivo, and found a slight decrease in fluorescence intensity in both the striatum and the hippocampus over time. However, the SNR remained largely stable in both brain regions (Extended Data Fig. 8). In the hippocampus, firing rates remained constant over time, and we were able to continuously image for up to 80 s with minimal changes in SNR (Extended Data Fig. 8i–m, Supplementary Video 1). This lack of toxicity is consistent with our results from cultured neurons (Extended Data Fig. 3).

Owing to the high performance and soma-targeted nature of SomArchon, we were able to routinely image multiple neurons at once in both cortical and subcortical brain regions (Figs. 2k, l, 3b, c, Extended Data Fig. 6a, b). In the hippocampus, using a 20× objective lens, we were able to record from 14 neurons at once, 8 of which were spiking (Fig. 4e, f). In addition, using a 16× objective lens, we routinely recorded from around 13 cells at once ($n = 13.1 \pm 3.5$ (mean ± s.d.) neurons per FOV from 13 recording sessions in 2 awake mice). Of the 170 manually identified neurons, 107 (63%) spiked during the recording periods (duration: 13.5–27 s) (Extended Data Fig. 6g–j, Supplementary Table 4). The ability to record from multiple neurons simultaneously enabled us to examine the correlation and coherence of subthreshold activities between pairs of neurons, although background fluorescence crosstalk between nearby neurons will need to

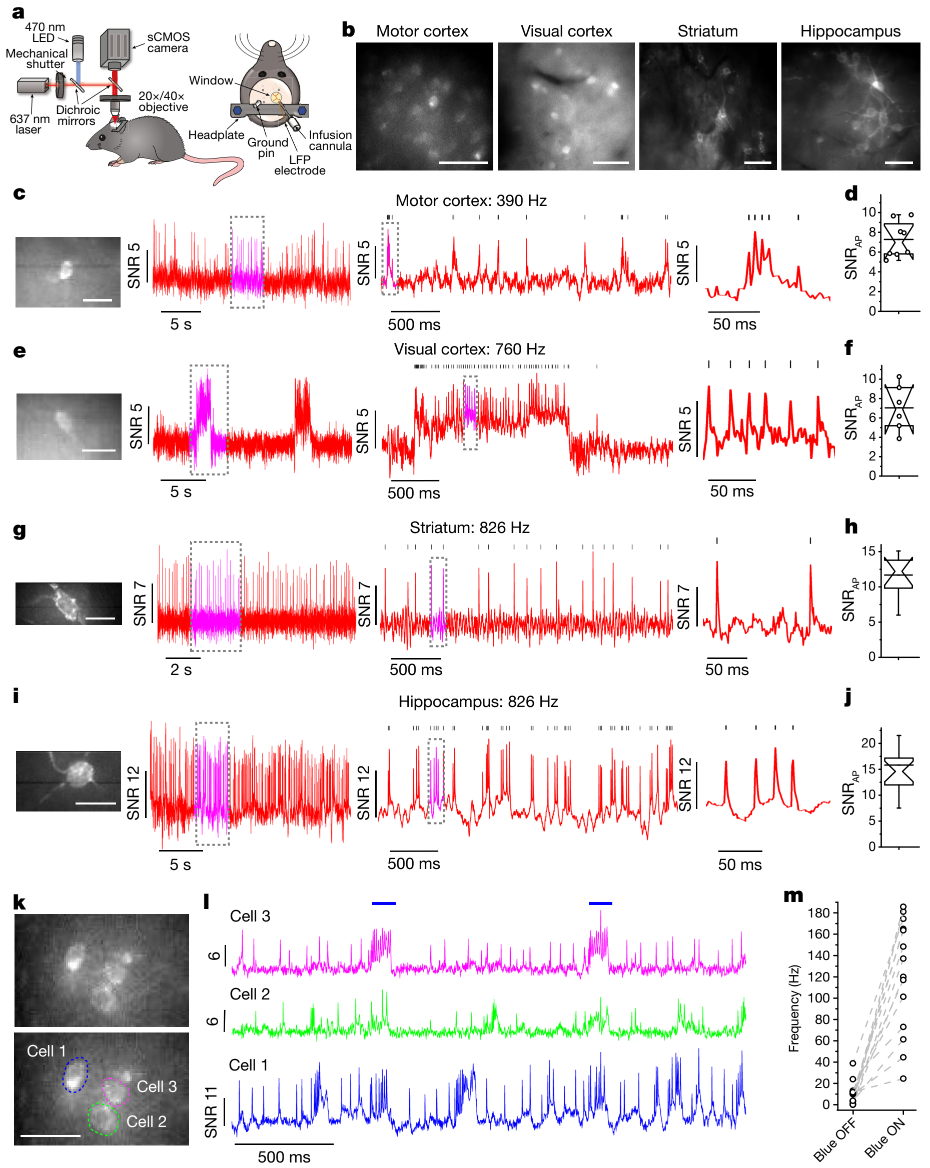


**Fig. 2 | SomArchon enables single-cell voltage imaging in multiple brain regions of awake mice, using a simple wide-field imaging setup.** **a**, Experimental setup. Left, awake mice had their heads fixed under a wide-field microscope; right, surgical window implant coupled with an infusion cannula and an LFP recording electrode. **b**, Representative SomArchon-expressing neurons visualized via EGFP fluorescence in motor cortex, visual cortex, striatum, and hippocampus ($\lambda_{ex} = 470/25$ nm LED, $\lambda_{em} = 525/50$ nm). Scale bars, 50 μm. **c**, **e**, **g**, **i**, Voltage imaging in motor cortex (**c**), visual cortex (**e**), striatum (**g**), and hippocampus (**i**). Left, SomArchon fluorescence image of the cell in vivo; right, optical voltage trace acquired from the cell (dashed boxes indicate time intervals shown at successively expanded time scales; vertical bars indicate peaks of action potentials identified by a custom spike-sorting algorithm). $\lambda_{ex} = 637$ nm laser at 1.6 W mm$^{-2}$ for visual cortex and motor cortex, 4 W mm$^{-2}$ for striatum and hippocampus, $\lambda_{em} = 664$ long-pass. Scale bars, 25 μm. **d**, **f**, **h**, **j**, Quantification of SNR per action potential for motor cortex (**d**), visual cortex (**f**), striatum (**h**), and hippocampus (**j**). Box plots as in Fig. 1. In **b**–**j**, representative images and traces were selected from, and statistics performed on, $n = 8$, 6, 10, and 17 cells from 3, 2, 3, and 4 mice for the motor cortex, visual cortex, striatum, and hippocampus, respectively. **k**, Fluorescence image of selected FOV showing hippocampal neurons expressing SomArchon–P2A–CoChR–K$_V$2.1$_{motif}$ (top) with neurons identified (bottom); $\lambda_{ex} = 637$ nm, exposure time 1.2 ms. Scale bar, 20 μm. **l**, Representative single-trial optical voltage traces from cells shown in **k** with blue light stimulation (100 ms pulse). Image acquisition rate, 826 Hz. **m**, Firing rate changes during blue light off versus blue light on conditions in individual neurons. In **k**–**m**, representative image selected from, and statistics performed on, $n = 14$ cells from 2 mice.

be considered when interpreting pairwise correlation and coherence measurements (Extended Data Figs. 9e–h, 10, Supplementary Discussion).

Compared to existing fully genetically encoded voltage indicators, SomArchon achieves a severalfold improvement in the number of cells that can be imaged simultaneously, while using inexpensive one-photon microscopy. The previously published record for fully genetically encoded voltage imaging was four spiking cells recorded simultaneously in an awake behaving mouse, but this required a specialized imaging setup that combined two-photon structural imaging to support patterned single-photon excitation illumination targeting individual cell bodies, as well as the blue-light-gated molecule paQuasAr3, which is incompatible with commonly used pulsed blue light optogenetic modulation[4]. ASAP3 has been used to image three dendrites at once, with 2-photon microscopy, but this approach can be used to record only single cells at the fast rates typical for voltage imaging[5]. The hybrid GEVI sensor Voltron enables imaging of 46 neurons[6], but requires the addition of chemicals delivered to the living brain that complicate in vivo mammalian use, is not compatible with optogenetics, and exhibits a lower dynamic range than SomArchon (Extended Data Fig. 5). Voltron, ASAP3, and Ace2N-mNeon all exhibit crosstalk with rhodopsins, hampering their use with optogenetic actuators. Voltage

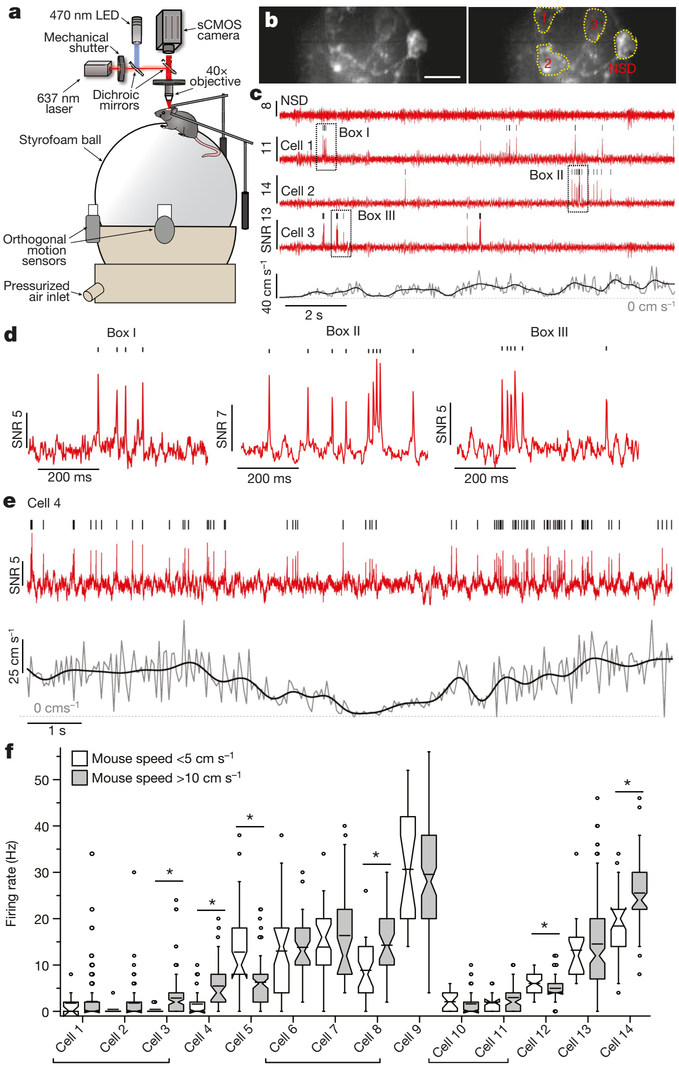


**Fig. 3 | Voltage imaging of striatal neurons during locomotion. a**, Schematic of the experimental setup, similar to that in Fig. 2a, but with mice positioned on a spherical treadmill. Imaging was performed with a 40× objective lens. **b**, Left, SomArchon fluorescence image of striatal cells; right, identified regions of interest (ROIs) corresponding to somas. $\lambda_{ex} = 637$ nm, exposure time 1.2 ms (representative image selected from $n = 9$ FOV from 2 mice). Scale bar, 20 μm. NSD, no spikes detected. **c**, Optical voltage traces acquired from cells in **b** and corresponding mouse movement speed (black, low-pass-filtered at 1.5 Hz; grey, raw data; representative traces selected from $n = 2$ FOV from one mouse). Image acquisition rate, 826 Hz. **d**, Magnified views of the three periods indicated by black boxes in **c**. **e**, Optical voltage trace (red) for a neuron modulated by movement speed and corresponding movement speed (black and grey; representative trace selected from $n = 14$ neurons from 2 mice). **f**, Firing rates of individual striatal neurons, during periods with low (open box plots) versus high (grey box plots) movement speed ($n = 14$ neurons from 2 mice, brackets indicate neurons from the same FOV). $^{*}P < 0.05$, two-sided Wilcoxon rank-sum test. Box plots as in Fig. 1.

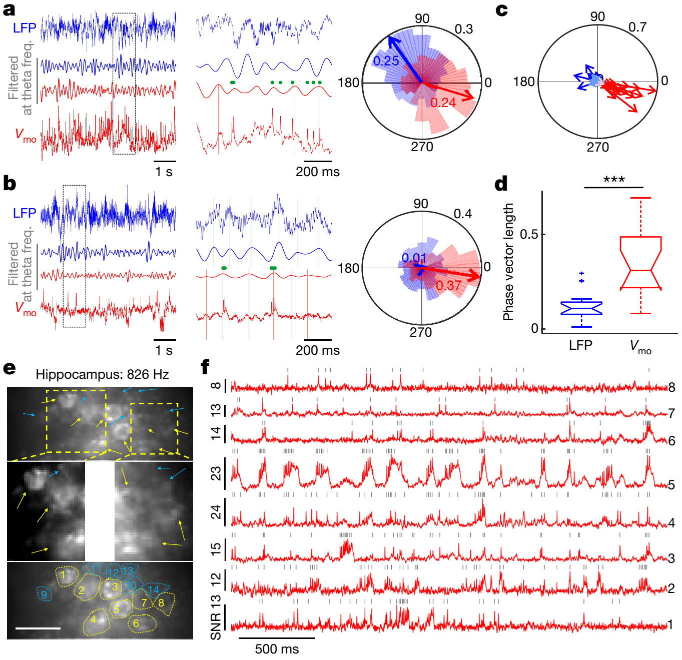


**Fig. 4 | Population voltage imaging of spikes and subthreshold voltage activities in CA1 neurons. a**, Neuron with spikes phase-locked to theta oscillations of LFPs (blue) and optically recorded membrane voltage ($V_{mo}$, red). Left, raw LFPs (top) and $V_{mo}$ (bottom), and theta frequency-filtered traces (middle). Middle, magnified view of the boxed period on the left. Theta oscillation peaks are indicated by blue and red vertical lines, and spikes by green dots. Right, probability distribution of the timing of spikes relative to the phases of the LFP (blue) and $V_{mo}$ (red) oscillations at theta frequency. Arrows indicate the average phase vector (with vector length indicated). Outer circle number indicates probability. Example selected from $n = 16$ neurons in 7 FOVs from 4 mice. **b**, As in **a**, but for an example neuron phase-locked to $V_{mo}$ theta oscillations, but not to LFP theta oscillations. **c**, Population spike–phase vectors relative to theta oscillations of LFP (blue and light blue) and $V_{mo}$ (red and pink). Each vector represents the average vector from one neuron (blue and red, $P < 0.05$; light blue and pink, not significant; $\chi^2$ test, spike–phase distribution of each neuron against uniform distribution; $n = 198$–1,077 spikes per neuron; 16 neurons in 7 FOVs from 4 mice). **d**, Population spike–phase relationship. $^{***}P = 5.0 \times 10^{-5}$, two-tailed paired Student's $t$-test, $n = 16$ neurons in 7 FOVs from 4 mice. Box plots: 25th and 75th percentiles with notch, median; whiskers, all data points not considered outliers; plus, outliers. **e**, SomArchon fluorescence images of CA1 neurons (top) with ROIs overlaid (bottom; $n = 14$ FOV from 3 mice). Middle, zoomed-in views of the yellow boxes from the top. Yellow arrows, example spiking cells with optical voltage traces shown in **f**; blue arrows, neurons not active during the period shown. $\lambda_{ex} = 637$ nm laser at 1.5 W mm$^{-2}$. Scale bar, 20 μm.

imaging with SomArchon or QuasAr3 is limited mainly by the high power and illumination spot of the 637-nm excitation laser; however, our data suggest that high-powered 637-nm excitation does not induce more phototoxicity than is seen with common, lower-powered 470-nm excitation (Extended Data Fig. 3). In conclusion, SomArchon is fully genetically encoded, is compatible with conventional easily accessible one-photon wide-field fluorescence microscopes, is fully compatible with blue-light-driven optogenetics, and enables routine imaging of around 13 neurons in a single FOV. We anticipate that the practicality of SomArchon will enable its rapid deployment into a variety of contexts in neuroscience. As camera performance improves in years to come, and as further evolution of GEVIs continues, we anticipate that it might be possible to image tens to hundreds of neurons using simple one-photon optics in the near future.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1641-1.

Received: 23 August 2018; Accepted: 28 August 2019;
Published online 9 October 2019.

1. Hochbaum, D. R. et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **11**, 825–833 (2014).
2. St-Pierre, F. et al. High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor. *Nat. Neurosci.* **17**, 884–889 (2014).

3. Gong, Y., Wagner, M. J., Zhong Li, J. & Schnitzer, M. J. Imaging neural spiking in brain tissue using FRET-opsin protein voltage sensors. *Nat. Commun*. **5**, 3674 (2014).
4. Adam, Y. et al. Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. *Nature* **569**, 413–417 (2019).
5. Chavarha, M. et al. Fast two-photon volumetric imaging of an improved voltage indicator reveals electrical activity in deeply located neurons in the awake brain. Preprint at https://www.biorxiv.org/content/10.1101/445064v2 (2018).
6. Abdelfattah, A. S. et al. Bright and photostable chemigenetic indicators for extended *in vivo* voltage imaging. *Science* **365**, 699–704 (2019).
7. Gong, Y. et al. High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor. *Science* **350**, 1361–1366 (2015).
8. Lou, S. et al. Genetically targeted all-optical electrophysiology with a transgenic Cre-dependent optopatch mouse. *J. Neurosci*. **36**, 11059–11073 (2016).
9. Jin, X., Tecuapetla, F. & Costa, R. M. Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nat. Neurosci*. **17**, 423–430 (2014).
10. Shi, L. H., Luo, F., Woodward, D. J. & Chang, J. Y. Neural responses in multiple basal ganglia regions during spontaneous and treadmill locomotion tasks in rats. *Exp. Brain Res*. **157**, 303–314 (2004).
11. Gritton, H. J. et al. Unique contributions of parvalbumin and cholinergic interneurons in organizing striatal networks during movement. *Nat. Neurosci*. **22**, 586–597 (2019).
12. Flytzanis, N. C. et al. Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and *Caenorhabditis elegans* neurons. *Nat. Commun*. **5**, 4894 (2014).
13. Piatkevich, K. D. et al. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol*. **14**, 352–360 (2018).
14. Jin, L. et al. Single action potentials and subthreshold electrical events imaged in neurons with a fluorescent protein voltage probe. *Neuron* **75**, 779–785 (2012).
15. Chamberland, S. et al. Fast two-photon imaging of subcellular voltage dynamics in neuronal tissue with genetically encoded indicators. *eLife* **6**, e25690 (2017).
16. Zou, P. et al. Bright and fast multicoloured voltage reporters via electrochromic FRET. *Nat. Commun*. **5**, 4625 (2014).
17. Gong, Y. et al. High-speed recording of neural spikes in awake mice and flies with a fluorescent voltage sensor. *Science* **350**, 1361–1366 (2015).
18. Shemesh, O. A. et al. Temporally precise single-cell-resolution optogenetics. *Nat. Neurosci*. **20**, 1796–1806 (2017).
19. Baker, C. A., Elyada, Y. M., Parra, A. & Bolton, M. M. L. Cellular resolution circuit mapping with temporal-focused excitation of soma-targeted channelrhodopsin. *eLife* **5**, e14193 (2016).
20. Daigle, T. L. et al. A suite of transgenic driver and reporter mouse lines with enhanced brain-cell-type targeting and functionality. *Cell* **174**, 465–480.e22 (2018).
21. Wu, C., Ivanova, E., Zhang, Y. & Pan, Z. H. rAAV-mediated subcellular targeting of optogenetic tools in retinal ganglion cells *in vivo*. *PLoS ONE* **8**, e66332 (2013).
22. Klapoetke, N. C. et al. Independent optical excitation of distinct neural populations. *Nat. Methods* **11**, 338–346 (2014).
23. Kravitz, A. V. & Kreitzer, A. C. Striatal mechanisms underlying movement, reinforcement, and punishment. *Physiology (Bethesda)* **27**, 167–177 (2012).
24. Koós, T. & Tepper, J. M. Inhibitory control of neostriatal projection neurons by GABAergic interneurons. *Nat. Neurosci*. **2**, 467–472 (1999).
25. Zhou, F. M., Wilson, C. J. & Dani, J. A. Cholinergic interneuron characteristics and nicotinic properties in the striatum. *J. Neurobiol*. **53**, 590–605 (2002).
26. Bittner, K. C. et al. Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat. Neurosci*. **18**, 1133–1142 (2015).
27. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).

## METHODS

**Molecular cloning.** To screen candidates for the soma-localized Archon1 voltage sensor in primary hippocampal neurons, we synthesized DNAs coding for candidate localization motifs de novo with mammalian codon optimization and subcloned them with the genes for Archon1 (GenBank ID MG250280.1) and EGFP into the pAAV-CAG vector to obtain the final constructs (Supplementary Table 1; gene synthesis and subcloning performed by Epoch Life Science). For in vivo expression in the mouse brain via IUE, the genes encoding Archon1–KGC–EGFP–$K_V2.1_{motif}$–ER2, QuasAr3–PP–Citrine–$K_V2.1_{motif}$–ER2 (QuasAr3-s), paQuasAr3–PP–Citrine–$K_V2.1_{motif}$–ER2 (paQuasAr3-s) and CoChR–mTagBFP2–$K_V2.2_{motif}$–ER2 were subcloned into the pCAG-WPRE vector. The genes encoding QuasAr3–PP–Citrine–$K_V2.1$–ER2 and paQuasAr3–PP–Citrine–$K_V2.1$–ER2 were synthesized de novo (GenScript Biotech) based on the reported sequences[28]. The gene for CoChR–mTagBFP2–$K_V2.2_{motif}$–ER2 was assembled by Epoch Life Science using pAAV-Syn–CoChR–GFP (Addgene plasmid no. 59070) and pBAD–mTagBFP2 (Addgene plasmid no. 34632) as the source of the genes for CoChR and mTagBFP2, respectively; the $K_V2.2$ motifs were synthesized de novo with mammalian codon optimization (Epoch Life Science). The pAAV-Syn-Archon1-KGC-EGFP-$K_V2.1_{motif}$-P2A-CoChR-$K_V2.1_{motif}$ plasmid was also cloned by Epoch Life Science. We used the Kv2.1 motif fused to CoChR for the following reason: in our original paper on soma-targeted CoChR[18], we used the KA2 sequence, which worked best with CoChR–GFP, but in this paper we used fluorophore-free CoChR, which did not express well with KA2, and rather worked better with $K_V2.1$; we also sometimes used the corresponding sequence from $K_V2.2$ as described in the text. Plasmid amplification was performed using Stellar (Clontech Laboratories) or NEB10-beta (New England BioLabs) chemically competent *Escherichia coli* cells. Small-scale isolation of plasmid DNA was performed with Mini-Prep kits (Qiagen); large-scale DNA plasmid purification was done with GenElute HP Endotoxin-Free Plasmid Maxiprep Kits (Sigma-Aldrich). The genes for ASAP3-Kv and Voltron-ST were synthesized de novo by GenScript, based on the reported sequences[5,6], and cloned into the pCAG-WPRE vector.

**Neuronal culture and transfection.** All mouse procedures were performed in accordance with the National Institute of Health Guide for Laboratory Animals and approved by the Massachusetts Institute of Technology Institutional Animal Care and Use and Biosafety Committees. For preparation of dissociated hippocampal mouse neuron cultures, we used postnatal day 0 or 1 Swiss Webster mice without regard to sex (Taconic Biosciences) as previously described[13]. In brief, dissected hippocampal tissue was digested with 50 units of papain (Worthington Biochemical) for 6–8 min at 37 °C, and the digestion was stopped by incubation with ovomucoid trypsin inhibitor (Worthington Biochemical) for 4 min at 37 °C. The tissue was then gently dissociated with Pasteur pipettes, and dissociated neurons were plated at a density of 20,000–30,000 per glass coverslip coated with Matrigel (BD Biosciences). Neurons were seeded in 100 μl plating medium containing MEM (Life Technologies), glucose (33 mM, Sigma), transferrin (0.01%, Sigma), HEPES (10mM, Sigma), Glutagro (2 mM, Corning), insulin (0.13%, Millipore), B27 supplement (2%, Gibco), and heat-inactivated FBS (7.5%, Corning). After cell adhesion, additional plating medium was added. AraC (0.002 mM, Sigma) was added when glia density was 50–70% of confluence. Neurons were grown at 37 °C and 5% $CO_2$ in a humidified atmosphere.

For in vitro screening of candidate soma-localized Archon1 sensors, primary hippocampal neuron cultures were transfected with 500 ng plasmid DNA per well using a commercial calcium phosphate transfection kit (Life Technologies) after 4 days in vitro (DIV), as previously described[13]. After 30–60 min of incubation of cultured neurons with DNA-calcium phosphate precipitate at 37 °C, neurons were washed twice with acidic MEM buffer (pH 6.7–6.8) to remove residual calcium phosphate particles and returned to the original plating medium. All measurements on cultured neurons were taken between DIV 14 and DIV 18 (about 9–14 d post transfection) to allow sodium channel maturation (and thus spiking). No cultured neuron recordings were supplemented with all-*trans*-retinal.

**Electrophysiology and fluorescence microscopy in cultured primary hippocampal neurons.** Whole-cell patch-clamp recordings of cultured neurons for Supplementary Table 1 were acquired via an Axopatch 700B amplifier (Molecular Devices) and Digidata 1440 digitizer (Molecular Devices). Neurons were patched between DIV 14 and DIV 18 and were bathed in Tyrode's solution (125 mM NaCl, 2 mM KCl, 3 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES, 30 mM glucose, pH 7.3 (NaOH adjusted)) at 32 °C during measurements. Synaptic blockers (2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo[f]quinoxaline (NBQX), 10 μM; D(−)-2-amino-5-phosphonovaleric acid, 25 μM; gabazine, 20 μM; Tocris) were added to the extracellular solution for single-cell electrophysiology. Borosilicate glass pipettes with an outer diameter of 1 mm and a wall thickness of 0.2 mm were pulled to produce electrodes with resistance of 3–10 MΩ and were filled with an internal solution containing 135 mM potassium gluconate, 8 mM NaCl, 10 mM HEPES, 4 mM Mg-ATP, 0.4 mM Na-GTP, 0.6 mM $MgCl_2$, 0.1 mM $CaCl_2$, pH 7.25 (KOH adjusted) at 295 mOsm. Measurements from primary neuron cultures were performed on the electrophysiology setup described above. Patch-clamp data were acquired only if the resting potential was below −45 mV and access resistance was <25 MΩ. Access resistance was compensated at 30–70%. Fluorescence imaging was performed on an inverted fluorescence microscope (Nikon Ti), equipped with a red laser (637 nm, 100 mW, Coherent, OBIS 637LX, Pigtailed) expanded by a beam expander (Thorlabs) and focused onto the back focal plane of a 40× NA 1.15 objective lens (Nikon).

Two-photon imaging of SomArchon-expressing neurons was performed using an Olympus FVMPE-RS equipped with two lasers for fluorescence excitation. An InSight X3 laser (Spectra-Physics) tuned to 1,150 nm at 50% transmissivity was used to excite SomArchon, and a MaiTai HP Ti:Sapphire laser (Spectra-Physics) tuned to 920 nm at 15% transmissivity was used to excite EGFP. The laser beams were focused using a 25× 1.05 NA water-immersion objective lens (Olympus). SomArchon emission was separated using a 590-nm dichroic mirror and imaged with 660–750 nm and 495–540 nm filters for near-infrared and green fluorescence, respectively, and signals were collected onto separate photomultiplier tubes. Imaging was performed at 2.0 μs per pixel sampling speed with one-way galvano scanning.

**Phototoxicity and photobleaching measurements in cultured neurons.** For phototoxicity and photostability measurements, primary mouse neuron cultures, prepared as described above, were imaged using an inverted Eclipse Ti-E microscope (Nikon) equipped with an sCMOS camera (OrcaFlash4.2, Hamamatsu), LED light source (Spectra, Lumencor), a 637-nm laser (637 LX, OBIS) focused on the back focal plane of a 40× NA 1.15 water immersion objective lens (Nikon), and a Polygon400 Multi-wavelength Patterned Illuminator (Mightex) with a 470-nm LED (ThorLabs). To express SomArchon, neurons were infected with AAV2-CaMKII-SomArchon or AAV2-Syn-SomArchon-P2A-CoChR-$K_V2.1_{motif}$ at DIV 5. To express ASAP3-Kv and Voltron-ST, neurons were transfected with the pCAG-ASAP3-Kv-WPRE and pCAG-Voltron-ST-WPRE plasmids, respectively, using the calcium phosphate method described above. For imaging of Voltron-expressing neurons, the cells were incubated with JF525 at final concentration 1.25 μM for 60 min at 37 °C (application of higher concentrations of JF525 resulted in marked internalization of the dye within 40 min of incubation at 37 °C, thus preventing functional imaging owing to the high background fluorescence). After incubation, the cells were washed 3 times with fresh plating medium for 3 h to removed unbound dye. The reactive oxygen species (ROS) measurements were performed using CellRox Orange dye (Invitrogen) according to the manufacturer's protocol. In brief, neurons were incubated with the CellROX Orange reagent at a final concentration of 5 μM for 30 min at 37 °C in darkness, and then washed once with fresh plating medium before imaging. Immediately before imaging, cells were supplemented with the NucGreen Dead 488 reagent for detection of plasma membrane integrity, which we used to indicate cell death. Cells that showed a more than ten times increase in green fluorescence in the nucleus over background fluorescence levels were considered dead. Neurons were imaged between DIV 14 and DIV 18 in the plating medium at 22 °C. CellROX Orange fluorescence was acquired using 510/25 nm excitation at 0.8 $mW/mm^2$ and 545/40 nm emission. NucGreen fluorescence was acquired using 475/36 nm excitation at 3.5 $mW/mm^2$ and 527/50 nm emission.

**IUE, AAV injection, and acute brain slice preparation.** For IUE, embryonic day (E)15.5 timed-pregnant female Swiss Webster (Taconic Biosciences) mice were deeply anaesthetized with 2% isoflurane. Uterine horns were exposed and periodically rinsed with warm sterile PBS. Plasmid DNA (1–2 μg total at a final concentration of about 2–3 μg/μl diluted in sterile PBS) was injected into the lateral ventricle of one cerebral hemisphere of an embryo. Five voltage pulses (50 V, 50 ms duration, 1 Hz) were delivered using 5-mm round plate electrodes (ECM 830 electroporator, Harvard Apparatus), with the anode or cathode placed on top of the skull to target the cortex or hippocampus, respectively. Electroporated embryos were placed back into the dam, and allowed to mature to delivery. Brain slices were prepared from electroporated mice without regard to sex at postnatal day (P)12–P22.

The electroporated mice were anaesthetized by isoflurane inhalation and decapitated, and cerebral hemispheres were quickly removed and placed in cold choline-based cutting solution consisting of (in mM): 110 choline chloride, 25 $NaHCO_3$, 2.5 KCl, 7 $MgCl_2$, 0.5 $CaCl_2$, 1.25 $NaH_2PO_4$, 25 glucose, 11.6 ascorbic acid, and 3.1 pyruvic acid (339–341 mOsm/kg; pH 7.75 adjusted with NaOH) for 2 min, then blocked and transferred into a slicing chamber containing ice-cold choline-based cutting solution. For mice electroporated with Voltron-ST, 50 μl JF525 dye (Janelia Farm; 12.5 nM JF525 in 10μl DMSO mixed with 10 μl Pluronic F-127 (20% w/v in DMSO; Invitrogen) and 30 μl sterile PBS) was injected into the retro-orbital sinus one day before slicing. Coronal slices (300 μm thick) were cut with a Compresstome VF-300 slicing machine, then transferred to a holding chamber with artificial cerebrospinal fluid (ACSF) containing (in mM) 125 NaCl, 2.5 KCl, 25 $NaHCO_3$, 2 $CaCl_2$, 1 $MgCl_2$, 1.25 $NaH_2PO_4$ and 11 glucose (300–310 mOsm/kg; pH 7.35 adjusted with NaOH), and allowed to recover for 10 min at 34 °C, followed by another 30 min at room temperature. Slices were subsequently maintained at room temperature (22 °C) until use. Both cutting solution and ACSF were constantly bubbled with 95% $O_2$ and 5% $CO_2$.

For AAV injection, 21-day-old C57 BL/6J mice were anaesthetized with isoflurane and placed in a small animal stereotaxic apparatus (David Kopf Instruments). Animals were injected with 200 nl rAAV8-Syn-Archon1-KGC-EGFP-Kv2.1motif-ER2 using a Nanoject (Drummond Scientific) via glass pipettes with 20–30-μm diameter tips into the striatum: anteroposterior (AP) 1.2 mm, mediolateral (ML) 2.1 mm, dorsoventral (DV) 3.2 mm relative to bregma. Brain slices were then prepared from these AAV-injected mice at P30–35. Mice were deeply anaesthetized with isoflurane and perfused transcardially using cold saline containing (in mM): 194 sucrose, 30 NaCl, 4.5 KCl, 1.2 $NaH_2PO_4$, 0.2 $CaCl_2$, 2 $MgCl_2$, 26 $NaHCO_3$, and 10 D-(+)-glucose saturated with 95% $O_2$ and 5% $CO_2$, pH 7.4 adjusted with NaOH, 320–340 mOsm/l. Coronal slices (250–300 μm thick) were cut using a slicer (VT1200 S, Leica Microsystems) and then incubated for 10–15 min in a holding chamber (BSK4, Scientific System Design) at 32 °C with regular ACSF containing (in mM): 136 NaCl, 3.5 KCl, 1 $MgCl_2$, 2.5 $CaCl_2$, 26 $NaHCO_3$ and 11 glucose saturated with 95% $O_2$ and 5% $CO_2$, followed by at least 1 h recovery at room temperature (21–25 °C) before recording.

**Concurrent electrophysiology and fluorescence imaging in acute brain slices.** For the recordings shown in Fig. 1 and Extended Data Fig. 3a–c, individual slices were transferred to a recording chamber mounted on an upright microscope (Olympus BX51WI, see below) and continuously superfused (2–3 ml/min) with carbogenated ACSF at room temperature. Whole-cell patch-clamp recordings were performed with borosilicate glass pipettes (KG33, King Precision Glass) heat-polished to obtain direct current resistances of ~4–6 MΩ. For cortex recordings, pipettes were filled with an internal solution containing in mM: 120 K-gluconate, 2 $MgCl_2$, 10 HEPES, 0.5 EGTA, 0.2 $Na_2ATP$, and 0.2 $Na_3GTP$. For hippocampus and striatum recordings, pipettes were filled with an internal solution containing in mM: 131 K-gluconate, 17.5 KCl, 9 NaCl, 1 $MgCl_2$, 10 HEPES, 1.1 EGTA, 2 $Na_2ATP$, and 0.2 $Na_3GTP$. Voltage clamp recordings were made with a microelectrode amplifier (Multiclamp 700B, Molecular Devices). Cell membrane potential was held at −60 mV, unless specified otherwise. Signals were low-pass-filtered at 2 kHz and sampled at 10–20 kHz with a Digidata 1440A (Molecular Devices), and data were stored on a computer for subsequent offline analysis. Cells in which the series resistance ($R_s$, typically 8–12 MΩ) changed by >20% were excluded from subsequent data analysis. In addition, cells with $R_s$ more than 25 MΩ at any time during the recordings were discarded. In some cases, conventional characterization of neurons was made in both voltage and current clamp configurations. Positive neurons were identified for recordings on the basis of EGFP expression visualized with a microscope equipped with a standard GFP filter (BX-51WI, Olympus). Optical voltage recordings were taken through a 40× water immersion objective (Olympus LUMFL N 40x/0.8W). Fluorescence was excited using a fibre-coupled 637-nm red laser (140 mW, Coherent Obis 637-140 LX), and the emission was filtered through a 664-nm long-pass filter. Images were collected on an EMCCD camera (Andor iXON Ultra 888) or sCMOS camera (Andor Zyla4.2 Plus Andor) in a reduced pixel window to enable acquisition at about 1 kHz. Each trial was about 30 s in duration. Of the 18 cortical neurons reported in Fig. 1, 4 neurons were not analysed for Fig. 1e because the electrophysiology files were inadvertently not saved, owing to a problem with the manual save process.

For optical recordings shown in Fig. 1i, j and Extended Data Figs. 1a–g, 2, 5, acute brain slices were transferred to a recording chamber mounted on an inverted Eclipse Ti-E (Nikon) equipped with a CMOS camera (Zyla5.5, Andor), LEDs (Spectra, Lumencor), a 637-nm laser (637 LX, OBIS) focused on the back focal plane of a 40× NA 1.15 objective (Nikon), and a Polygon400 Multiwavelength Patterned Illuminator (Mightex) with 470-nm LED (ThorLabs), and continuously superfused (2–3 ml/min) with carbogenated ACSF at room temperature. Positive cells were imaged under 0.8 or 1.5 W/mm$^2$ (55 mW) excitation light power at 637 nm from the laser. 4-Aminopyridine at a final concentration of 1 mM was added to induce neuronal activity for experiments shown in Extended Data Figs. 2c, d, g, h, 5. For Fig. 1i, j and Extended Data Fig. 5, cells were illuminated with 2-ms blue light pulses at light power in the range from 0.1 to 1.0 mW/mm$^2$.

**Mouse surgery.** All in vivo mouse procedures were performed in accordance with the National Institute of Health Guide for Laboratory Animals and approved by the Boston University Institutional Animal Care and Use and Biosafety Committees.

*Virus injection surgery.* All AAVs were produced by the University of North Carolina Chapel Hill Vector Core. Adult female C57BL/6 mice (Charles River Laboratories) or Chat-cre mice (Chat-cre;129S6-Chattm2(cre)Lowl/J, the Jackson Laboratory), 8–12 weeks old at the time of surgery, were used for all experiments. AAV-Syn-SomArchon ($5.9 \times 10^{-12}$ genome copies (GC)/ml) or AAV-syn-SomArchon-P2A-CoChR-Kv2.1 ($2.19 \times 10^{-13}$ GC/ml) was injected into the motor cortex (AP: +1.5 mm, ML: ±1.5 mm, DV: −0.3 mm relative to bregma, 0.5 μl virus), visual cortex (AP: −3.6 mm, ML: ±2.5 mm, DV: −0.3 mm, 0.5 μl virus), hippocampus (AP: −2.0 mm, ML: +1.4 mm, DV: −1.6 mm, 1 μl virus) or striatum (AP: +0.8 mm, ML: −1.8 mm, DV: −2.1 mm, 1 μl virus). Viral injection occurred at 50–100 nl/min (10 min total) using a 10-μl syringe (NANOFIL, World Precision Instruments) fitted with a 33-gauge needle (World Precision Instruments, NF33BL) and controlled by a microinfusion pump (World Precision Instruments, UltraMicroPump3–4). The syringe was left in place for an additional 10 min after injection to facilitate viral spread. About one week after the viral injection, mice underwent a second surgery to implant the cranial window for in vivo imaging.

*Cortex imaging window implantation.* The imaging window consisted of a stainless steel cannula (OD: 3.17 mm, ID: 2.36 mm, height: 1 mm, AmazonSupply, B004TUE45E) fitted with a circular coverslip (no. 0, OD: 3 mm, Deckgläser Cover Glasses, Warner Instruments, 64-0726 (CS-3R-0)) adhered using a UV curable glue (Norland Products, Norland Optical Adhesive 60, P/N 6001). A craniotomy of about 3 mm in diameter was created, with the dura left intact, over the motor cortex (centred at AP: +1.5 mm, ML: ±1.75 mm) or visual cortex (AP: −3.6 mm, ML: ±2.15 mm). The imaging window was positioned over the cortex so that it was flush with the surface of the dura. Kwik-sil adhesive (World Precision Instruments, KWIK-SIL) was applied around the edges of the imaging window to hold the imaging window in place and to prevent any dental cement from touching the brain. Three small screws (J.I. Morris, F000CE094) were screwed into the skull to further anchor the imaging window to the skull. Dental cement was then gently applied to affix the imaging window to the exposed skull, and to mount an aluminium headbar posterior to the imaging window. Supplementary Fig. 1a, b provides the window placement.

*Hippocampus and striatum imaging window implantation.* Hippocampal and striatal window surgeries were performed in a similar way to those previously described[11,29]. For each imaging window, a virus/drug infusion cannula (26G, PlasticsOne, C135GS-4/SPC) was attached to a stainless steel imaging cannula (OD: 3.17 mm, ID: 2.36 mm, height: 1 or 2 mm, AmazonSupply, B004TUE45E). The bottom of the infusion cannula was flush with the base of the stainless steel cannula, and a circular coverslip (no. 0, OD: 3mm, Deckgläser Cover Glasses, Warner Instruments, 64-0726 (CS-3R-0)) was adhered using a UV curable glue (Norland Products, Norland Optical Adhesive 60, P/N 6001). An additional insulated stainless steel wire (diameter: 130 μm, PlasticsOne, 005SW-30S, 7N003736501F) was glued to the viral/drug infusion cannula with super glue (Henkel, Loctite 414 and Loctite 713) and protruded from the bottom of the infusion cannula and imaging window by about 200 μm for LFP recordings.

A craniotomy about 3 mm in diameter was made over the hippocampus CA1 region (AP: −2.0 mm, ML: +2.0 mm) or the striatum (AP: +0.8 mm, ML: −1.8 mm). A small notch was made on the posterior edge of the craniotomy to accommodate the infusion cannula and LFP recording electrode. The overlying cortex was gently aspirated using the corpus callosum as a landmark. The corpus callosum was then carefully thinned to expose the hippocampus CA1 region or the dorsal striatum. The imaging window was positioned in the craniotomy, and Kwik-sil adhesive (World Precision Instruments, KWIK-SIL) was applied around the edges of the imaging window to hold it in place and to prevent any dental cement from touching the brain. Three small screws (J.I. Morris, F000CE094) were screwed into the skull to further anchor the imaging window to the skull, and a small ground pin was inserted into the posterior part of the brain near the lambda suture as a ground reference for LFP recordings. Dental cement was then gently applied to affix the imaging window to the exposed skull, and to mount an aluminium headbar posterior to the imaging window. Supplementary Fig. 1c, d provides the window placement.

In mice that did not receive a virus injection before window implantation, 1 μl of AAV-syn-SomArchon ($5.9 \times 10^{12}$ GC/ml) or 1 μl of AAV-syn-SomArchon-P2A-CoChR-Kv2.1 ($2.19 \times 10^{13}$ GC/ml), or 1 μl of AAV-CAG-FLEX-SomArchon ($6.3 \times 10^{12}$ GC/ml) was injected through the virus/drug infusion cannula at 100 nl/min through an internal infusion cannula (33G, PlasticsOne, C315IS-4/SPC) connected to a microinfusion pump (World Precision Instruments, UltraMicroPump3–4), one week after the window implantation surgery. The internal infusion cannula was left in place for 10 min after injection to facilitate viral spread. Mice were awake and their heads were fixed throughout the injection period.

All mice were treated with buprenex for 48 h after surgery and single-housed to prevent any damage to the headbar or window implant.

**In vivo imaging in the live mouse brain.** All optical recordings were acquired on a conventional one-photon fluorescence microscope equipped with an ORCA Flash 4.0 V3 Digital CMOS camera (Hamamatsu Photonics K.K., C13440-20CU) or Hamamatsu ORCA Fusion Digital CMOS camera (Hamamatsu Photonics K.K., C14440-20UP), 10× NA0.25 LMPlanFI air objective (Olympus), 40× NA0.8 LUMPlanFI/IR water immersion objective (Olympus), 20× NA1.0 XLUMPlanFL N water immersion objective (Olympus), 16× NA0.8 CFI LWD Plan Fluorite water immersion objective (Nikon), 470-nm LED (ThorLabs, M470L3), 140-mW 637-nm red laser (Coherent Obis 637-140X), a green filter set with a 470/25-nm bandpass excitation filter, a 495-nm dichroic, and a 525/50-nm bandpass emission filter, and a near infrared filter set with a 635-nm laser dichroic filter, and a 664-nm long-pass emission filter. The near-infrared laser illuminated a circular area of about 60–80 μm, about 80–140 μm, and about 100–200 μm in diameter in brain tissue, with FOV height (limited by camera acquisition rate) 40–60 μm, 80–100 μm, and 100–120 μm under 40×, 20×, and 16× objective lenses, respectively. A mechanical shutter (Newport, model 76995) was positioned in the laser path to control the timing of illumination over the imaging window. Optical recordings were acquired at 390–900 Hz with HCImage Live (Hamamatsu Photonics K.K.) or NIS Elements (Nikon) software. HC Image Live data were stored as DCAM image files (DCIMG), and further analysed offline in Fiji/ImageJ and MATLAB (Mathworks). NIS Elements data were stored as .nd2 files and further analysed offline using the NIS Elements software.

The GFP signal of SomArchon was acquired in the green channel ($\lambda_{ex} = 470$ nm) at 1,024 × 1,024 pixels with 2 × 2 binning to show cell structure and distribution. Optical voltage recordings were imaged in the near infrared channel ($\lambda_{ex} = 637$ nm) with 2 × 2 or 4 × 4 binning. OmniPlex system (PLEXON) was used to synchronize data acquisition from different systems. In all experiments, the OmniPlex system recorded the start of image acquisition from the sCMOS camera, the acquisition time of each frame, and other experiment-dependent signals described below.

*Optical imaging of spontaneous neural activity.* All in vivo optical imaging of spontaneous neural activity was performed when mice were awake with their heads fixed in a custom holder that allowed attachment of the headplate at the anterior end. Animals were covered with an elastic wrap to prevent upward movement. Spontaneous neural activity recordings lasted continuously for up to 30,000 frames (about 36 s).

*Eye puff.* During some in vivo hippocampal imaging recordings, an eye puff was applied to evoke high-frequency local field potential responses in the hippocampus (Extended Data Fig. 9a–d). The mice had their heads fixed in a custom holder that allowed attachment of the headplate at the anterior end, and they were covered with an elastic wrap to prevent upward movement. Each experimental session consisted of 20–30 trials, with each trial lasting for 5,000 frames (about 6 s). Three seconds after the start of image acquisition, the sCMOS camera sent a TTL pulse to a function generator (Agilent Technologies, model 33210A), which triggered a 100-ms-long air puff. The air puff was 5–10 psi, and administered via a 0.5-mm cannula placed 2–3 cm away from the eye of the mouse. The puff TTL pulses were also recorded with the OmniPlex system (PLEXON). Eye movement was monitored using a USB webcam (Logitech, Carl Zeiss Tessar 2.0/3.7 2MP Autofocus).

*Optopatch blue light stimulation.* All in vivo optopatch (that is, optogenetics plus voltage imaging) experiments were performed when mice were awake with their heads fixed in a custom holder that allowed attachment of the headplate at the anterior end. Mice were covered with an elastic wrap to prevent upward movement. A 470-nm LED (ThorLabs, M470L3) was coupled to a Polygon400 Multiwavelength Patterned Illuminator (Mightex), and the blue light was focused through the objective lens to illuminate the centre of the FOV. At the onset of imaging, the sCMOS camera sent a TTL pulse to trigger Axon CNS (Molecular Devices, Digidata 1440A) which controlled the 470-nm LED (ThorLabs). Each trial lasted 1.1 s and consisted of a single 100-ms-long blue light pulse, 500 ms after trial onset. Each recording session consisted of 10 trials with increasing blue light power from 0.1 to 1 mW/mm$^2$, with a step of about 0.1 mW/mm$^2$ per trial. The OmniPlex system (PLEXON) recorded the timing of TTL pulses used to trigger the Axon CNS.

*Head-fixed voluntary movement experiments.* All voluntary movement experiments were performed while awake, head-fixed mice were freely navigating a spherical treadmill. The spherical treadmill was constructed as described[30]. In brief, a 3D spherical Styrofoam ball was supported by air, and motion was tracked using two computer mouse sensors positioned roughly ±45° from the centre along the equator of the ball. All motion-sensor displacement data were acquired at 100 Hz on a separate computer and synthesized using a custom Python script. Motion sensor displacement data were then sent to the image acquisition computer to be accumulated using a modified ViRMEn MATLAB script. The timing of each motion sensor displacement data point was also recorded using the OmniPlex system (PLEXON) to synchronize movement data with optical voltage recordings.

To determine the mouse movement speed, ball movement was first calibrated. The ball was pinned on the two sides and rotated vertically to calibrate sensor displacement.

All mice were habituated on the spherical treadmill for at least 3 days, at least 20 min per day, before image acquisition. During optical imaging, mice were imaged while freely navigating the spherical treadmill. Each FOV was recorded for at least 36 s in total. In some fields of view, we performed multiple trials, and each trial was at least 12 s in duration with an inter-trial interval of at least 30 s in duration.

**Local field potential recording.** Local field potentials were recorded using an OmniPlex system (PLEXON) at a 1 kHz sampling rate. To synchronize optical recordings with LFP recordings, the camera sent out a TTL pulse to the OmniPlex system at the onset of imaging and after each acquired frame.

**Motion correction.** In Figs. 2i, 3, 4, motion correction was performed with a custom Python script. For FOV, if multiple video imaging files were collected for the same FOV, we started with the first imaging file to ensure speedy data processing (a single video file contains a series of images). We first generated the reference image by averaging across all images within the file. We then performed a series of image processing procedures to enhance the contrast of the reference image and every image in the file to facilitate motion correction. We first removed 10% of the pixels along all edges of an image to remove any camera induced artefact. We then applied a high-pass filter (Python scipy package, ndimage.gaussian_filter, sigma = 50) to remove low-frequency components within the images. To enhance the boundaries of high intensity areas, we identified the boundaries as the difference between two low-pass-filtered images (sigma = 2 and 1). We then enhanced the boundary by adding 100 times the boundary back to the low-pass-filtered image (sigma = 2). We then limited the intensity range of the processed images within one standard deviation above and below the average intensity of the image, by setting the pixels with intensity higher than mean + s.d. as mean + s.d., and the pixels with intensity lower than mean − s.d. as zero. Finally, to counter any potential bleaching over time, we normalized the intensity of each image by shifting the mean intensity to zero and divided intensity values by the s.d. of all pixel intensities in that image. After image processing, we calculated the displacement of each image, by identifying the maximum cross-correlation coefficient between each image and the reference image, and then corrected motion by shifting the displacement in the original, unenhanced image sequence. If the same FOV was imaged over an extended period of time, during which multiple files were acquired, we motion-corrected subsequent files by aligning them to the first file, so that the same ROIs from the same FOV could be applied across the entire imaging session. Specifically, we first refined the reference frame by generating the mean intensity projection image from the motion-corrected first imaging file. The refined reference image was then used to motion-correct all files of the same FOV, including the first file, using the procedure described above. The motion-corrected, original, unenhanced image sequences were then used for subsequent manual ROI segmentation and further analysis.

**ROI identification.** We imported the image files (motion-corrected as above, if necessary) into Fiji/ImageJ or NIS Elements and manually segmented ROIs by examining the time-series images to identify areas with clear neuron outlines and/or intensity dynamics over time. The optically recorded voltage traces for each ROI were generated from the motion-corrected image sequences using the multiple measurement function and were then used for analyses.

The wide dynamic range (16 bit) of the raw images meant that to select dim as well as bright cells, we had to create maximum projection and standard deviation images of the entire raw video, and stretched their look-up tables to enhance visibility. For Fig. 4 and Extended Data Fig. 9e–h, cells were densely packed, so we identified and tracked ROIs semi-manually across image sequences without performing motion correction. We first visually inspected all image sequences and identified those with minimal motion and with an SNR greater than about 2 for further analysis. We then performed an iterative ROI-selection procedure to identify ROIs that best fit each cell. Specifically, we started by manually selecting ROIs from the maximum projection image of the entire image sequence. The image sequence was then visually inspected to identify frames with cells that exhibited shifts of more than three pixels from the defined ROI. We then used these frames to separate the image sequence into multiple time intervals, and obtained a new set of maximum projection images to identify new ROIs within these time intervals for these cells. This procedure was repeated iteratively until the ROI represented the cell across all image frames in their corresponding time intervals without the cell moving out of the ROI. Thus, with this procedure, we created multiple ROIs representing the same cell across different frames. For each cell, we extracted traces for every ROI during its corresponding time interval, and stitched the baseline-normalized traces for the same cell(s) in time. The fluorescence traces of each cell were then detrended for further analysis. Supplementary Fig. 2 provides an example of raw and processed traces for two cells in the same FOV.

**Hippocampal spike detection.** Spikes were associated with a rapid increase in intensity, followed by a rapid decrease. By contrast, occasional motion artefacts were usually associated with a decrease in intensity as a neuron moved out of the ROI. To facilitate spike detection, we first removed motion artefacts. For each time point of the fluorescence intensity trace for each ROI, we calculated the change in intensity from that of the prior time point ($I_{change}$). We then defined noise as the time points at which instantaneous $I_{change}$ was 3 s.d. below the mean value of $I_{change}$ across the entire trace. We excluded any time points at which the $I_{change}$ of the previous time point was more than 1 s.d. above the average $I_{change}$, because this might have indicated a spike. These noise time points and their following three time points (as we found that motion artefacts are typically >4 ms) were then considered motion artefacts, and removed from further analysis. We then recalculated the standard deviation of $I_{change}$, excluding the data points related to the motion artefact. The peaks of spikes were then identified as time points that met the following two criteria: (1) the intensity change of the time point combined with that of its preceding time point was more than 3 s.d. above the average $I_{\text{change}}$, and (2) the intensity change over the next two time points was less than 2 s.d. below the average $I_{\text{change}}$.

**Hippocampal spike–phase calculation.** Hippocampal spike–phase analysis was performed on 16 neurons from 7 FOVs in 4 mice. For each FOV, we analysed data collected over 10 trials (about 60 s in total) during which animals experienced an eye puff in each trial, as described above. To calculate the phase of spikes at theta frequency (4–10 Hz), we first band-pass-filtered both the optical voltage trace and the simultaneously recorded LFP at theta frequency (eegfilt, EEGLAB toolbox). The peaks of theta oscillation power were then identified using the findspike function in MATLAB. For each spike, we obtained the phase of the spike by calculating the timing of each spike relative to the period of that oscillation cycle in degrees. We averaged the phases of all spikes from the same neuron as the average phase of a given neuron.

**Analysis for pairwise coherence between hippocampal neurons and LFPs.** The coherence analysis was performed on nine FOVs that contained multiple neurons from four mice. Each FOV contained imaging data over a period of 6–36 s. For each FOV, we first re-sampled the LFP at the acquisition rate of the optical imaging. We then divided the optical voltage traces and LFPs into segments of 1,000 data points. We then calculated the averaged coherence, at theta frequency (4–10 Hz), with the functions in the Chronux toolbox (optical voltage trace to optical voltage trace or LFP: coherencyc, and spike to spike: coherencypt) with tapers = [10 19], fpass = [4 10] and trialave = 1. To compare $V_{\text{mo}}$–$V_{\text{mo}}$ coherence with $V_{\text{mo}}$–LFP coherence across nine FOVs, we averaged the coherence of neurons in the same FOV to obtain the mean coherence of that FOV, and then performed statistical tests across FOVs using the mean coherences of the individual FOVs. To understand the relationships between pairs of coherence, we used the MATLAB function fitlm to perform a linear regression between coherent pairs and obtain the $P$ and $r^2$ values.

To estimate background fluorescence crosstalk, we calculated pairwise coherence and correlation between background doughnut areas surrounding a neuron. To select background doughnut areas, we excluded the edges (5%) of each FOV, because the edge may be missing for a particular image frame when image frames were shifted during motion correction. The background doughnut of a neuron was determined as the area 3–10 μm from the neuron boundary, excluding any pixels within 10 μm of the boundary of another neuron. One neuron was excluded from this analysis owing to dense labelling where we could not identify its doughnut area. Fluorescence traces of the background doughnut area were then processed as for neurons, and their pairwise coherence and Pearson correlations were calculated.

**Spike detection for striatum, motor cortex, and visual cortex.** After motion correction, we first identified large fluorescence increases using a threshold of 4 s.d. above the baseline. The baseline was manually selected as a period of >500 ms without spiking or drifting due to $z$-plane shifting or photobleaching. From these large fluorescence increases, we selected those with rise times and decay times shorter than 4 ms as spikes.

**Firing rate comparison of striatal neurons during high- and low-speed movement.** Movement data of mice were first interpolated to the voltage imaging frame rate with MATLAB function interp1, and then smoothed using a 1.5-Hz low-pass Butterworth filter to remove any motion sensor artefacts. We calculated the average movement speed at 0.5-s intervals and defined low-speed periods as intervals during which the average speed was ≤5 cm/s and high-speed periods as intervals during which the average speed was ≥10 cm/s. The firing rates during these high- and low-speed motion periods were compared, and a two-sided Wilcoxon rank-sum test was used to identify significant differences between these periods.

**SNR calculation for in vivo photostability evaluation over imaging duration in striatum and hippocampus.** We defined noise as the standard deviation of the fluorescence intensity across the entire trial period. For each neuron, we first calculated the SNR for each action potential by dividing the intensity change observed during an action potential by the noise, and then calculated the average SNRs across all spikes detected in a trial as the corresponding SNR for the trial. For the striatum dataset, only neurons imaged over at least five consecutive trials were analysed. For the hippocampus dataset, all neurons were analysed.

**Detrending.** All optically recorded SomArchon traces reported in the manuscript (except those shown in Fig. 4a–d) were corrected for photobleaching or focus shift by subtracting baseline fluorescence traces that were low-pass-filtered and fit to a double or single exponential function.

**Histology.** Mice were transcardially perfused with PBS followed by 4% paraformaldehyde. The brain was gently extracted from the skull and post-fixed in 4% paraformaldehyde for 1–4 h at room temperature or overnight at +4 °C. Fixed brains were transferred to a 30% sucrose–PBS solution and rotated for 24–48 h at 4 °C for cryoprotection. Cryoprotected brains were frozen in OCT in a dry ice bath and sliced (coronally) to 50-μm thickness using a cryostat. Glial and microglial antibody staining were performed with anti-GFAP[29] (1:250, Clone N206/A8, Neuromab) and anti-IBA1[31] (1:500, 019-19741, Wako Chemicals) primary antibodies, respectively, followed by Alexa Fluor 568 (1:1,000, goat anti-mouse IgG (H+L) cross-adsorbed secondary antibody, A11004, InVitrogen) and 633 secondary antibodies (1:1,000, goat anti-rabbit IgG (H+L) cross-adsorbed secondary antibody, A21070, InVitrogen). All antibodies were used according to protocols that have been validated by their suppliers. Slice imaging was performed using an inverted Nikon Eclipse Ti microscope equipped with a spinning disk sCSUW1 confocal scanner unit (Yokogawa), 488-, 561-, and 642-nm solid state lasers, 525/25-nm, 579/34-nm, and 664LP emission filters, a 20× NA0.75 air objective lens (Nikon), and a 4.2 PLUS Zyla camera (Andor), controlled by NIS-Elements AR software. Acquired images were contrast-enhanced to improve visualization.

**Brain temperature measurements.** Under general anaesthesia, a craniotomy about 3 mm in diameter was made to expose the brain surface, with a small notch on the posterior edge to accommodate the insertion of a temperature probe (Physitemp, IT-1E) coupled to a Thermocouple DAQ (DATAQ Instruments, Model DI-245). An imaging window, identical to those used in all imaging experiments, was positioned on the craniotomy. Kwik-sil adhesive was applied around the edges of the imaging window to hold it in place, but not around the craniotomy notch, to allow insertion of the temperature probe. Dental cement was then gently applied to affix the imaging window to the skull and to mount an aluminium headbar. Once mice had recovered from anaesthesia, they had their heads fixed while awake and the temperature probe was inserted under the imaging window above the brain surface, through the craniotomy notch. The 637-nm laser was directed through the 40× objective under identical conditions to those used while imaging (75–95 mW laser power), and brain temperature was recorded. We noted a temperature increase of 1.88 ± 0.80 °C (mean ± s.d., $n = 3$ mice) over the 12-s illumination period commonly used in our experiments. These changes are similar to, or smaller than, changes commonly seen with two-photon imaging, optogenetics, and the making of craniotomies for neural imaging[32–36].

**Sample size.** No statistical methods were used to estimate sample size for mouse studies throughout. We did not perform a power analysis, as our goal was to create a new technology, as recommended by the NIH: "In experiments based on the success or failure of a desired goal, the number of animals required is difficult to estimate..."[37]. As noted in the aforementioned paper, "The number of animals required is usually estimated by experience instead of by any formal statistical calculation, although the procedures will be terminated [when the goal is achieved]." These numbers reflect our past experience in developing neurotechnologies.

**Data exclusions.** Voltage imaging datasets with significant motion or in which no spikes were detected were excluded from analysis. Significant motion was defined as a shift of more than 20 μm in any direction. In Extended Data Fig. 1i–k, data points that corresponded to overlapping neurites were excluded. Data exclusion criteria were not pre-established.

**Replication.** All attempts at replication were successful.

**Randomization and blinding.** There were no treatment conditions to compare in this study. All recording sessions were randomly performed with different voltage sensors or in different brain regions. On recording days, cultured cells or brain slices expressing specific sensors were known. On in vivo recording days, mouse conditions were known. Voltage trace extraction and subsequent analysis were performed with the investigators unaware of specific mouse conditions. For analysis of movement modulation of striatal neuron spiking, a computer algorithm was used to identify periods with different movement parameters. For analysis of spike–phase relationships, or subthreshold membrane voltage relationships, a computer algorithm was used across all conditions. For histology, sections were selected and images were taken from slides by a researcher not aware of the conditions or antibody used. Cells were also counted and quantified from these sections by a researcher blinded to the experimental conditions or antibody used.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Code availability

Computer code used to generate results for this study is available at https://github.com/HanLabBU/somarchon-imaging.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request; raw data essential to the work are available online as Source Data files. Sequences of the reported proteins are available at GenBank at the following accession codes: SomArchon MN091368; SomArchon-P2A-CoChR-$K_V2.1_{\text{motif}}$, MN091369.

28. Adam, Y. et al. Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. *Nature* **569**, 413–417 (2019).
29. Hansen, K. R. et al. Mild blast injury produces acute changes in basal intracellular calcium levels and activity patterns in mouse hippocampal neurons. *J. Neurotrauma* **35**, 1523–1536 (2018).

30. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
31. Fields, I. E. et al. Noninvasive deep brain stimulation via temporally interfering electric fields. *Cell* **169**, 1029–1041.e16 (2017).
32. Kalmbach, A. S. & Waters, J. Brain surface temperature under a craniotomy. *J. Neurophysiol.* **108**, 3138–3146 (2012).
33. Podgorski, K. & Ranganathan, G. Brain heating induced by near-infrared lasers during multiphoton microscopy. *J. Neurophysiol.* **116**, 1012–1023 (2016).
34. Arias-Gil, G., Ohl, F. W., Takagaki, K. & Lippert, M. T. Measurement, modeling, and prediction of temperature rise due to optogenetic brain stimulation. *Neurophotonics* **3**, 045007 (2016).
35. Stujenske, J. M., Spellman, T. & Gordon, J. A. Modeling the spatiotemporal dynamics of light and heat propagation for in vivo optogenetics. *Cell Rep.* **12**, 525–534 (2015).
36. Christie, I. N. et al. fMRI response to blue light delivery in the naïve brain: implications for combined optogenetic fMRI studies. *Neuroimage* **66**, 634–641 (2013).
37. Dell, R. B., Holleran, S. & Ramakrishnan, R. Sample size determination. *ILAR J.* **43**, 207–213 (2002).

**Acknowledgements** We thank T. Ta for help with histology; M. Murdock for help with mouse work; N. Pak, A. Badon and J. Mertz for help with assembling the in vivo imaging setup; and Y. Bando for helping with the Supplementary Video. E.S.B. acknowledges funding from Edward and Kay Poitras, NIH Director's Pioneer Award 1DP1NS087724, NIH 1R01GM104948, NIH 1R01EB024261, NIH 1R01DA045549, NIH 1R01MH114031, Charles Hieken, NIH 1R01NS102727, John Doerr, NSF Grant 1734870, the HHMI-Simons Faculty Scholars Program, Human Frontier Science Program RGP0015/2016, NIH 1R43MH109332, US Army Research Laboratory and the US Army Research Office under contract/grant number W911NF1510548, NIH 2R01DA029639, and NSF CBET 1344219. X.H. acknowledges funding from the NIH Director's Office (1DP2NS082126), NINDS (1R01NS109794-01, 1R01NS087950-01), NSF CBET-1848029, the Grainger Foundation, the Pew Foundation, and Boston University Biomedical Engineering Department. S.B. and S.N.S. acknowledge funding from the NIH/NIGMS T32 Quantitative Biology and Physiology Fellowship (GM008764) through the Boston University Biomedical Engineering Department. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author contributions** K.D.P. and E.S.B. initiated the project. K.D.P., S.B., H.T., S.N.S., X.H., and E.S.B. designed all in vivo experiments and interpreted the data. K.D.P. developed SomArchon and together with E.E.J., O.A.S., and E.C. characterized all constructs in cultured cells. K.D.P., V.G.L.-H., D.P., C.S., Z.F., and B.L.S. performed characterization of SomArchon in acute brain slices. S.B., S.N.S. and H.J.G. performed all mouse surgeries for in vivo experiments. M.F.R. assisted on imaging setups. K.D.P, S.B., H.T., and S.N.S performed all in vivo imaging experiments and analysed all in vivo imaging data. K.D.P., S.B., H.T., S.N.S, X.H. and E.S.B. wrote the paper with contributions from all of the authors. E.S.B. and X.H. oversaw all aspects of the project.

**Competing interests** The authors declare no competing financial interests.

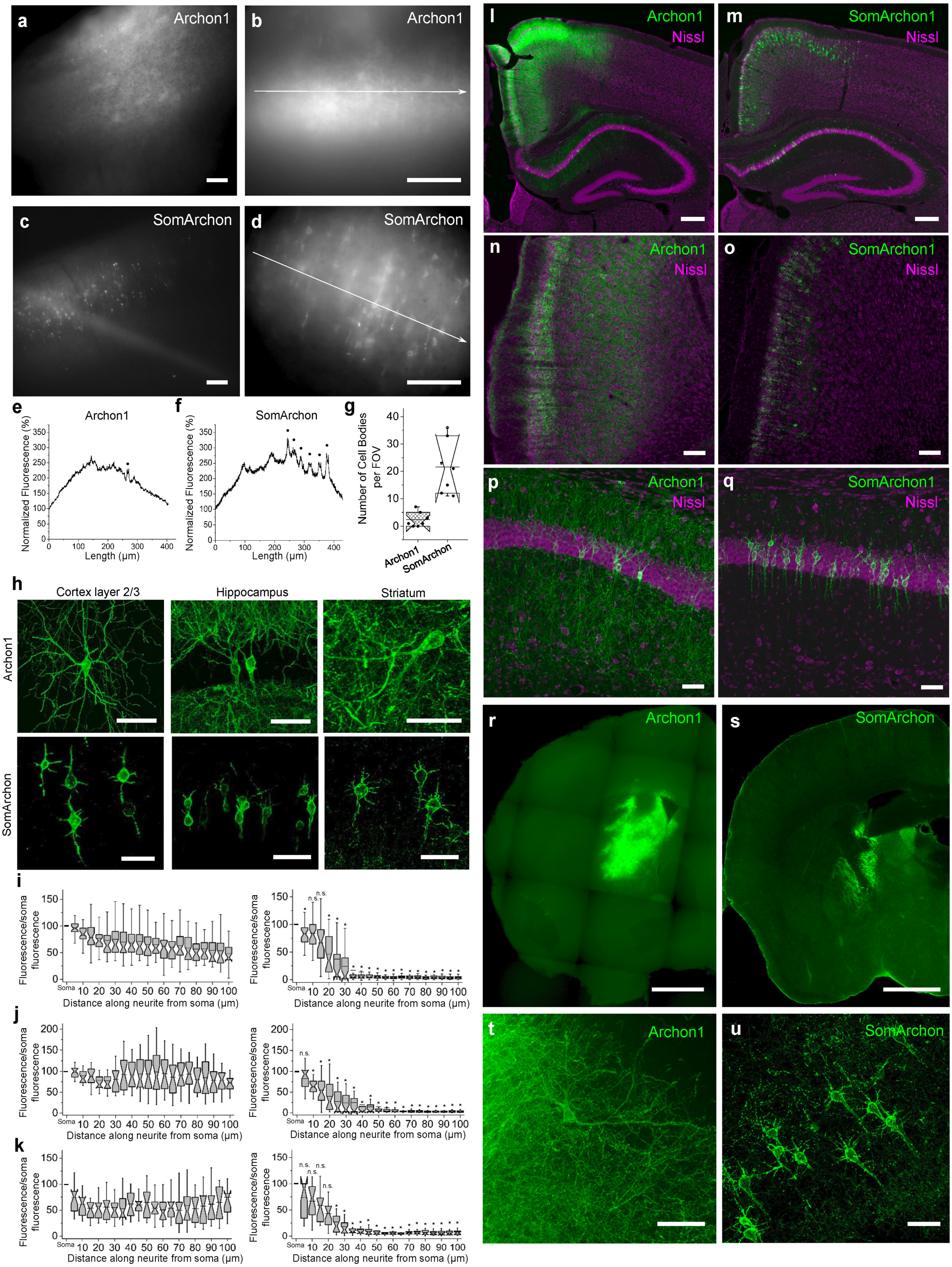
**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1641-1.

**Correspondence and requests for materials** should be addressed to E.S.B. or X.H.

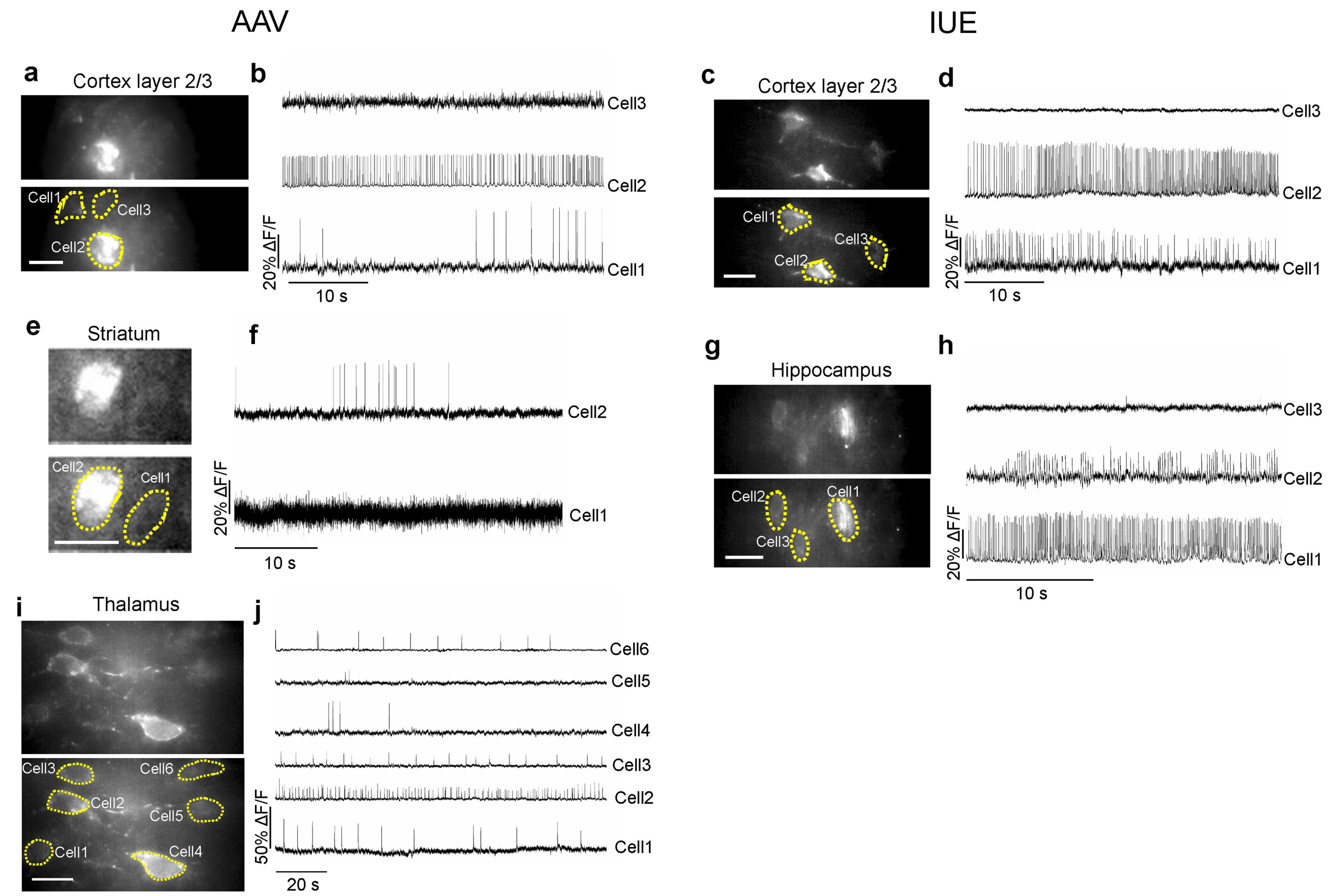
**Peer review information** *Nature* thanks Kenneth Harris, Lin Tian, Christian Wilms and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints.

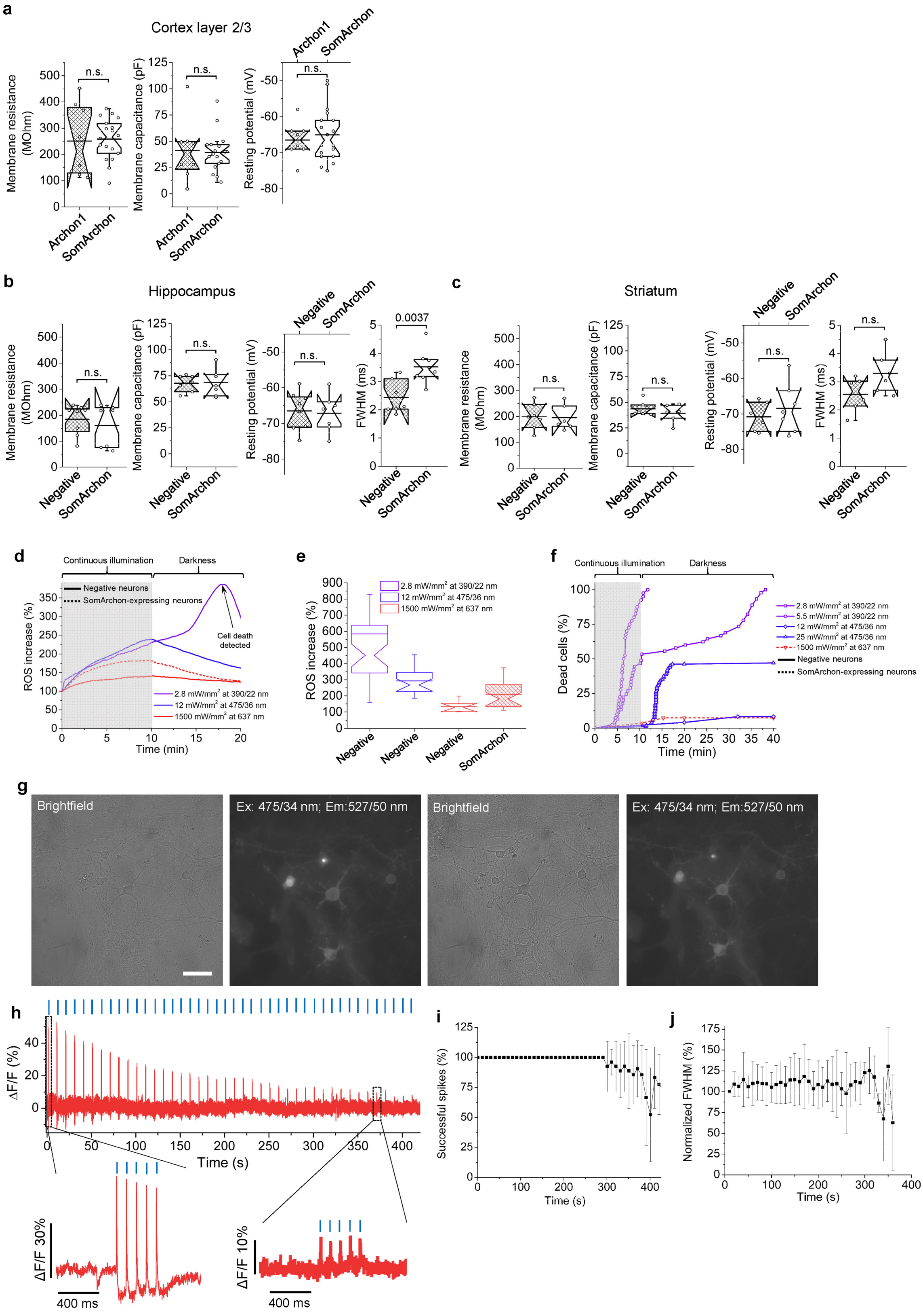


**Extended Data Fig. 1** | See next page for caption.

**Extended Data Fig. 1 | Expression of Archon1 and SomArchon in mouse brain. a**–**d**, Representative images of mouse brain slices expressing Archon1 (**a**, **b**) and SomArchon (**c**, **d**) (CAG promoter, via IUE) imaged with a wide-field microscope with 10× (**a**, **c**) and 40× (**b**, **d**) objective lenses (from $n = 7$ slices from 2 mice each). **e**, **f**, Normalized EGFP fluorescence along white arrows shown in **b**, **d**, respectively. Black dots correspond to resolvable cells. **g**, Number of resolvable cells per FOV for brain slices expressing Archon1 or SomArchon ($2.4 \pm 2.5$ and $22 \pm 9$ neurons per FOV ($350 \times 415\ \mu\text{m}^2$) for Archon1 and SomArchon, respectively). Mean ± s.d.; $n = 7$ slices from 2 mice each; box plots as in Fig. 1. Further confocal analysis with larger FOVs of $500 \times 500 \times 50\ \mu\text{m}^3$ revealed that SomArchon can resolve around 15 times more neurons in the cortex than Archon1 ($n = 4$, 8, 9, 11, 11, 18, and 20 neurons from 7 slices for Archon1, versus $n = 180$, 187, and 137 neurons from 3 slices for SomArchon). **h**, Representative confocal images of neurons in cortex layer 2/3 (left), hippocampus (middle), and striatum (right) expressing Archon1 (top) and SomArchon (bottom). **i**–**k**, EGFP fluorescence along a neurite, normalized to soma, for neurons expressing Archon1 (left) or SomArchon (right) in cortex layer 2/3 (**i**, $n = 39$ and 37 neurites from 10 cells from 2 mice each), hippocampus (**j**, $n = 20$ and 34 neurites from 9 and 17 cells from 2 mice each), and striatum (**k**, $n = 17$ and 20 neurites from 7 cells from 2 mice each). Box plots as in Fig. 1. $^*P < 0.002$ compared to Archon1 at corresponding position away from the soma; n.s., not significant. Two-sample Kolmogorov–Smirnov test, see Supplementary Table 2. **l**–**u**, Representative confocal fluorescence images of brain slices expressing Archon (left) or SomArchon (right) via IUE (**l**–**q**) or AAV injection (**r**–**u**) in cortex layer 2/3 (**n**, **o**; $n = 8$ slices from 2 mice), hippocampus (**p**, **q**; $n = 8$ slices from 2 mice), and striatum (**t**, **u**; $n = 6$ slices from 2 mice). Green, EGFP; magenta, Nissl staining. Scale bars, 100 μm (**a**–**d**), 50 μm (**h**, **n**–**q**, **t**, **u**), 250 μm (**l**, **m**, **r**, **s**).

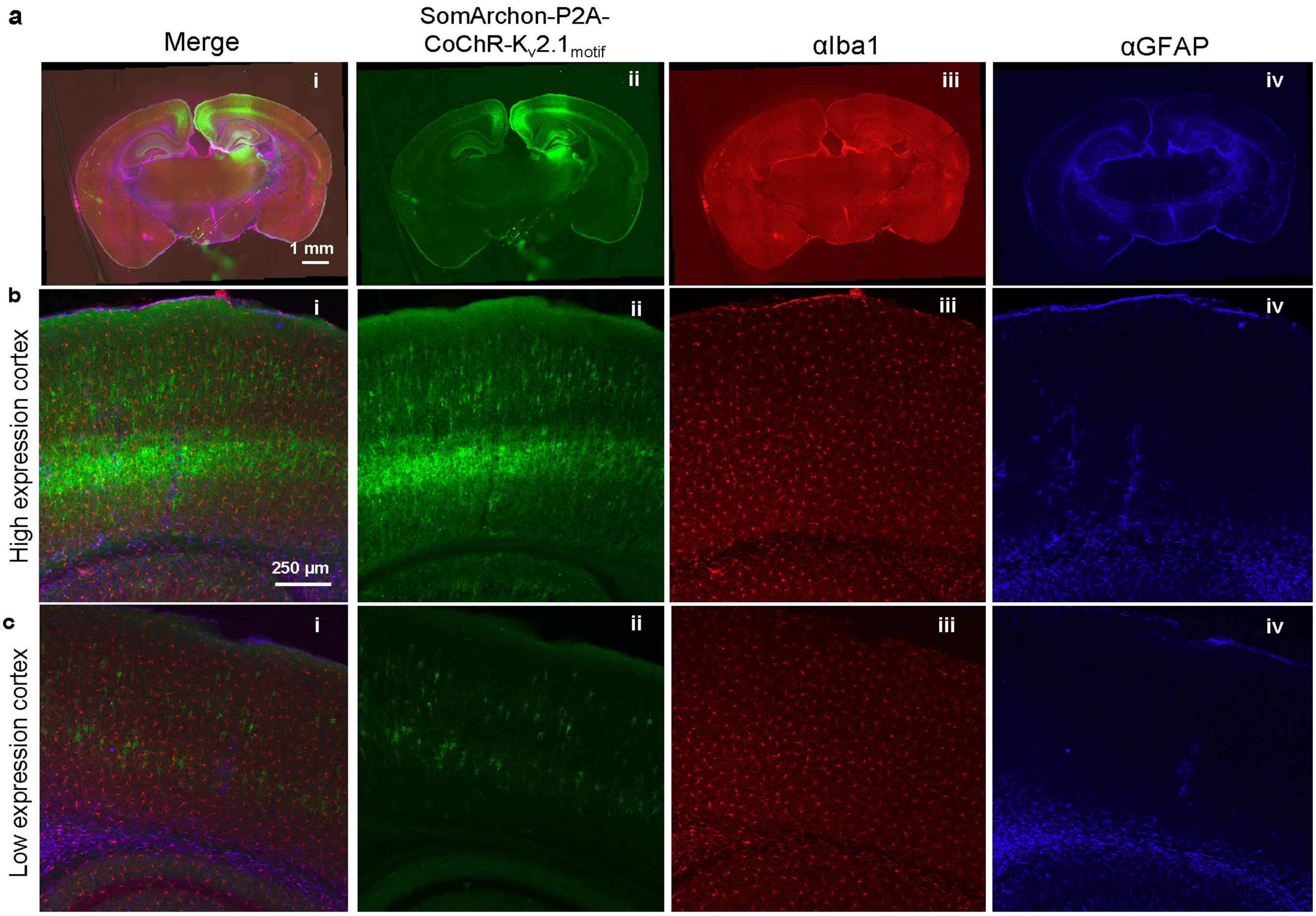


**Extended Data Fig. 2 | Voltage imaging using SomArchon in mouse brain slices. a–d**, Representative fluorescence wide-field images of cortex layer 2/3 neurons expressing SomArchon via AAV transduction (**a**) or IUE (**c**) with selected ROIs (bottom), and corresponding fluorescence traces (**b**, **d**; $n = 6$ and 13 slices from 2 and 4 mice for AAV transduction and IUE, respectively). Acquisition rate, 632 Hz (**b**) or 440 Hz (**d**). **e**, **f**, Representative fluorescence wide-field images of striatal neurons expressing SomArchon via AAV transduction (top) with selected ROIs (bottom) (**e**), and corresponding fluorescence traces (**f**; $n = 8$ slices from 2 mice). Acquisition rate, 733 Hz. **g**, **h**, Representative fluorescence wide-field images of hippocampal neurons expressing SomArchon via IUE (top) with selected ROIs (bottom) (**g**), and corresponding fluorescence traces (**h**; $n = 8$ slices from 2 mice). Acquisition rate, 333 Hz. **i**, **j**, Fluorescence wide-field images of thalamus neurons expressing SomArchon (top) via AAV transduction with selected ROIs (bottom) (**i**), and corresponding fluorescence traces (**j**; $n = 5$ slices from 2 mice). Acquisition rate, 333 Hz. Scale bars, 25 μm.

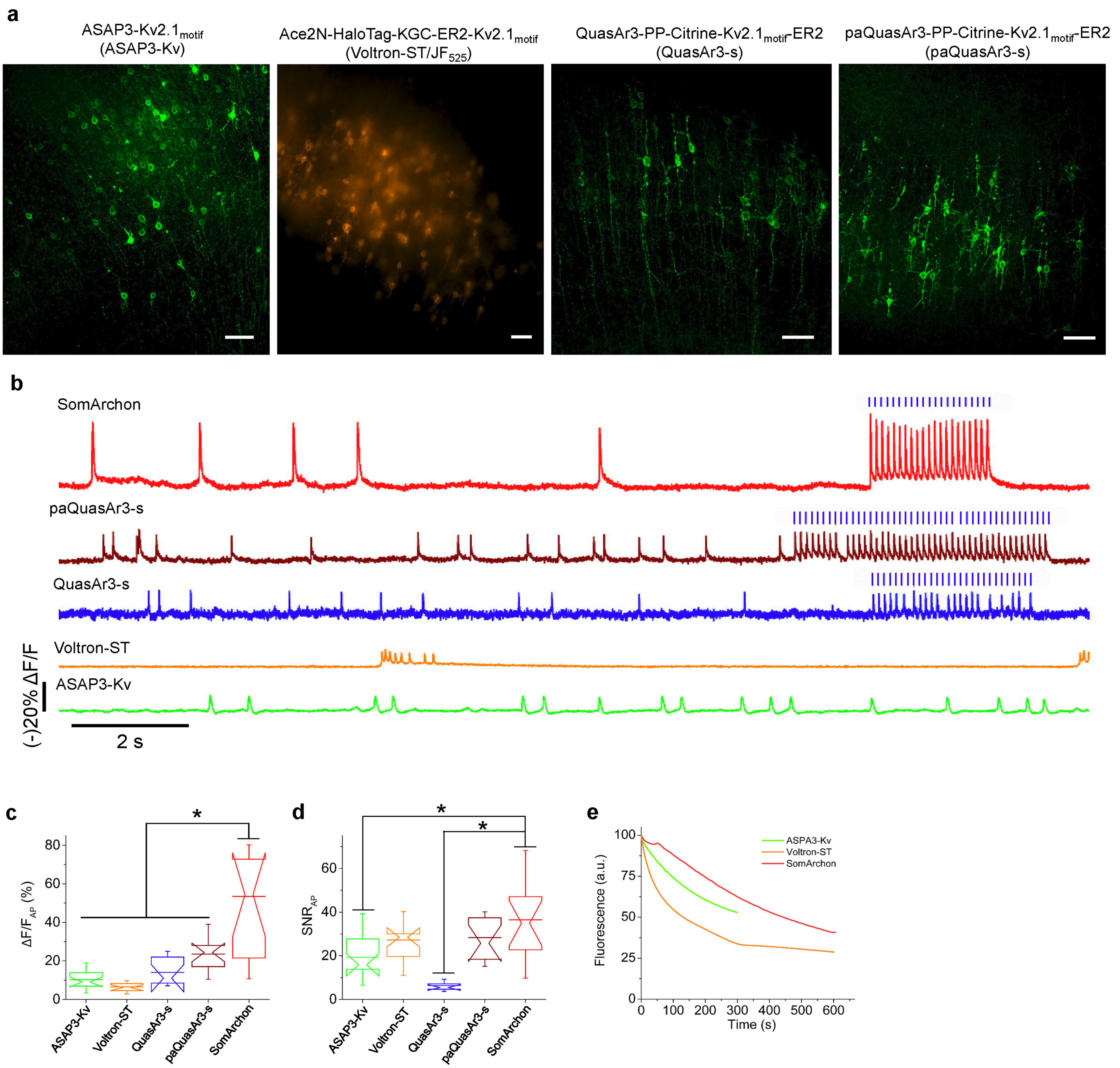


**Extended Data Fig. 3 |** See next page for caption.

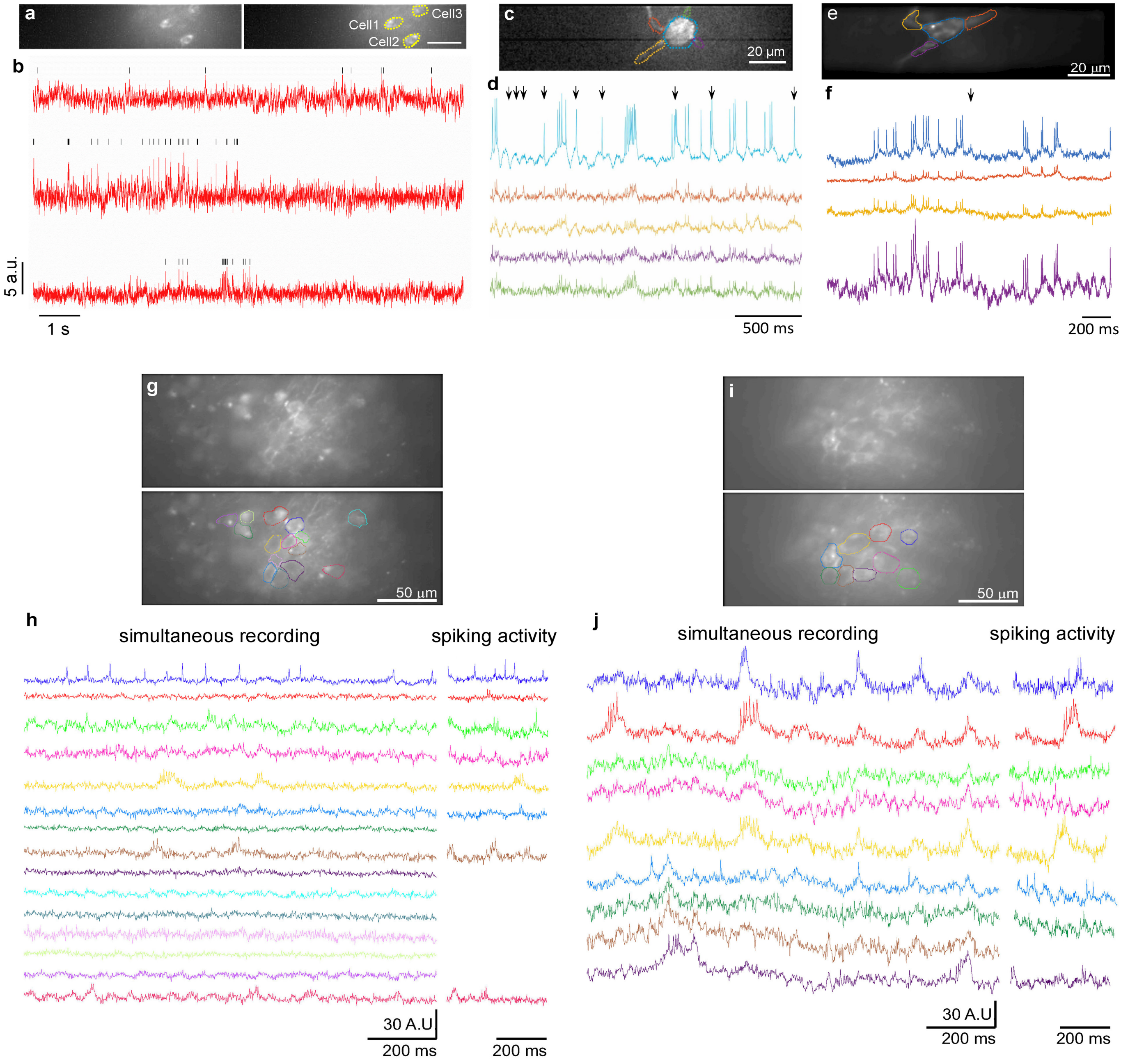
**Extended Data Fig. 3 | Expression of SomArchon and voltage imaging do not alter membrane properties or cause phototoxicity.** **a**, Membrane properties of neurons expressing Archon (hashed boxes) or SomArchon (open boxes) in cortex layer 2/3 brain slices ($P = 0.8026$, 0.8895, and 0.8236 for resistance, capacitance, and resting potential, respectively; two-sided Wilcoxon rank-sum test comparing Archon1 versus SomArchon; $n = 8$ and 18 cells from 1 and 2 mice for Archon1 and SomArchon, respectively). **b**, Similar to **a** but in hippocampus ($P = 0.6294$, 0.9720, 0.8880, and 0.0037 for resistance, capacitance, resting potential, and FWHM, respectively; two-sided Wilcoxon rank-sum test comparing negative versus SomArchon; $n = 8$ and 7 cells from 2 mice each for negative and SomArchon for resistance and resting potential; $n = 7$ and 7 cells from 2 mice each for negative and SomArchon for capacitance; $n = 7$ and 8 cells from from 2 mice each for negative and SomArchon for FWHM). **c**, Similar to **a** but in striatum ($P = 0.7380$, 0.8357, 0.7751, and 0.0931 for resistance, capacitance, resting potential and FWHM, respectively; two-sided Wilcoxon rank-sum test comparing negative and SomArchon; $n = 7$ and 6 cells from 2 mice each for negative and SomArchon for resistance and capacitance; $n = 6$ and 7 cells from 2 mice each for negative and SomArchon for resting potential; $n = 6$ and 6 cells from 2 mice each for negative and SomArchon for FWHM). **d**, Changes in relative ROS concentration (normalized to that before illumination) over time in negative (solid line) and SomArchon-expressing (dashed line) cultured mouse neurons under various illumination protocols. **e**, Maximal increase in ROS concentration during continuous illumination for conditions performed in **d** ($n = 45$, 24, and 8 negative neurons from 2, 2, and 1 cultures for 390/22 nm, 475/36 nm, and 637 nm illumination, respectively; $n = 24$ SomArchon-expressing neurons for 637 nm illumination from 1 culture). **f**, Cell death for negative (solid line) and SomArchon-expressing (dashed line) cultured neurons at DIV 14–18 under various illumination protocols ($n = 45$, 35, 91, 40, and 27 neurons from 2, 1, 2, 1, and 1 cultures, respectively, for 390/22 nm at 2.8 mW mm$^{-2}$, 390/22 nm at 5.5 mW mm$^{-2}$, 475/36 nm at 12 mW mm$^{-2}$, 475/36 nm at 25 mW mm$^{-2}$, and 637 nm at 1,500 mW mm$^{-2}$ illumination). **g**, Bright-field and fluorescence images of representative neurons expressing SomArchon before and after 10 min of continuous 637-nm laser illumination at 1,500 mW mm$^{-2}$, followed by 10 min in darkness (93% of imaged cells did not exhibit noticeable changes in morphology; $n = 27$ cells from 1 culture; non-illuminated cells did not show any changes in morphology; $n = 10$ cells from 1 cultures). Scale bar, 50 μm. **h**, Representative SomArchon fluorescence trace from neuron co-expressing SomArchon and CoChR-K$_V$2.1$_{\text{motif}}$. **i**, Normalized spike rates (to initial value) elicited by blue light illumination dropped after 300 s of continuous recording, owing to decrease in spike amplitude as a result of photobleaching ($n = 10$ neurons from 1 culture; plotted as mean ± s.d.). **j**, Normalized (to initial value) FWHM of spikes elicited by continuous light exposure as in **h**. Box plots as in Fig. 1.



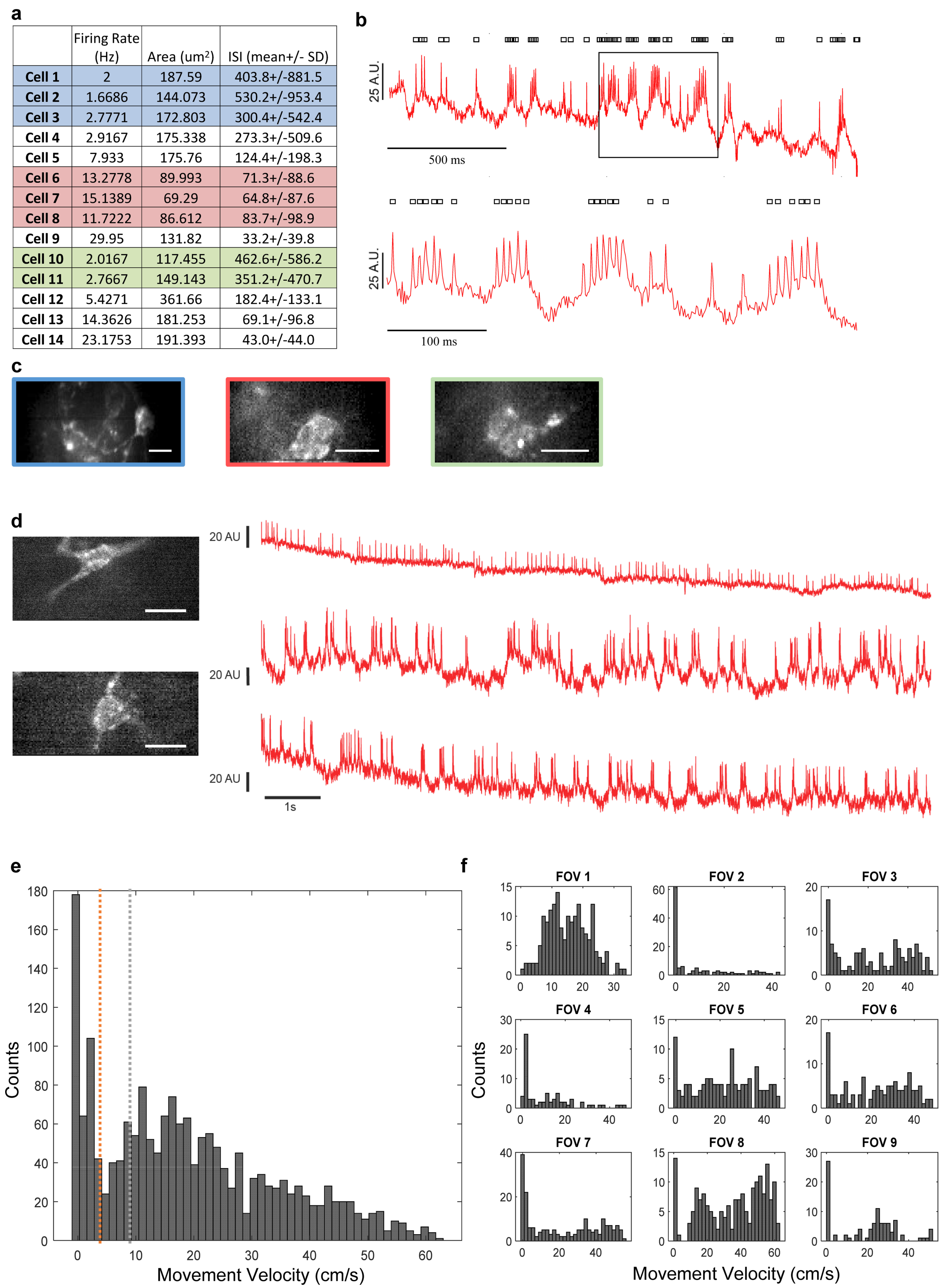
**Extended Data Fig. 4 | SomArchon expression in vivo does not cause gliosis.** SomArchon was expressed in the mouse brain by AAV2.9-Syn-SomArchon-P2A-CoChR-K$_V$2.1$_{motif}$ injection into the cortex in P0 Swiss Webster mice. Brain tissues were analysed 63 days after viral injection. Merged fluorescence images from 50-µm-thick coronal sections (i) were visualized via EGFP fluorescence of SomArchon (ii), anti-IBA1 immunofluorescence (iii), and anti-GFAP immunofluorescence (iv; $n = 4$ slices from 2 mice). **a**, Expression throughout the coronal section. **b**, Zoomed-in view of the virally injected area (high-expression cortex). **c**, Zoomed-in view of the non-injected contralateral hemisphere (low-expression cortex). The commonly used glial and microglial markers GFAP and IBA1 appeared similarly in both hemispheres, suggesting that expression of SomArchon did not cause gliosis. Scale bars, 1 mm (**a**), 250 µm (**b**, **c**).



**Extended Data Fig. 5 | Side-by-side comparison of next-generation voltage indicators in mouse brain slices.** **a**, Representative fluorescence images of mouse cortex layer 2/3 neurons expressing ASAP3-K$_V$2.1$_{motif}$ (ASAP3-Kv), Ace2N-HaloTag-KGC-ER2-K$_V$2.1$_{motif}$ (Voltron-ST/JF$_{525}$), QuasAr3-PP-mCitrine-K$_V$2.1$_{motif}$-ER2 (QuasAr3-s), and paQuasAr3-PP-mCitrine-K$_V$2.1$_{motif}$-ER2 (paQuasAr3-s). ASAP3-Kv, QuasAr3-s and paQuasAr3-s were visualized via cpGFP, mCitrine, and mCitrine fluorescence, respectively, using laser excitation at 488 nm and emission at 525/50 nm under a confocal microscope. Voltron-ST/JF$_{525}$ was visualized via JF525 fluorescence using LED excitation at 510/25 nm and emission at 545/40 nm under a wide-field microscope. Scale bar, 50 μm. **b**, Single-trial optical recordings of ASAP3-Kv (green) and Voltron-ST/JF$_{525}$ (orange) fluorescence responses during neuronal activity evoked with 4-aminopyridine, and QuasAr3-s (blue), paQuasAr3-s (brown), and SomArchon (red) fluorescence responses during neuronal activity evoked with CoChR-mTagBFP2-K$_V$2.2$_{motif}$. Acquisition rate, about 500 Hz. Blue light pulses (470/20 nm, 2 ms per pulse, 10 Hz; vertical blue bars) were used to activate CoChR to evoke spiking. **c**, **d**, Quantification of $\Delta F/F$ (**c**) and SNR (**d**) per action potential across all recordings ($n = 18$, 14, 9, 13, and 14 neurons from 1, 2, 2, 2, and 2 mice for ASAP3-Kv, Voltron-ST/JF$_{525}$, QuasAr3-s, paQuasAr3-s, and SomArchon, respectively). Box plots as in Fig. 1. $*P < 0.01$, Wilcoxon rank-sum test; see Supplementary Table 2 for statistics. **e**, Photobleaching curves of ASAP3-Kv, Voltron-ST/JF$_{525}$, and SomArchon under continuous illumination ($n = 11$, 8, and 17 neurons from 1 culture, respectively).



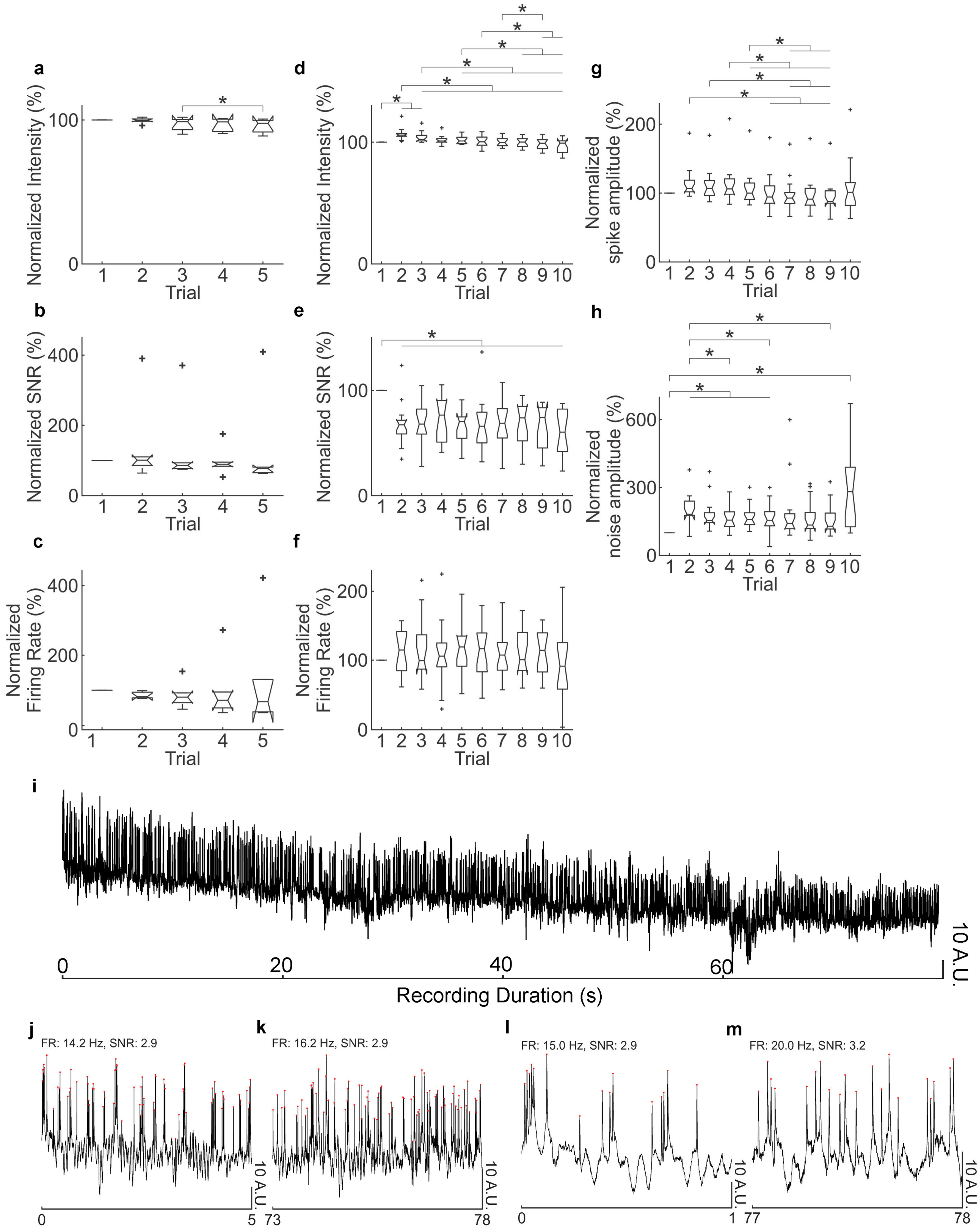
**Extended Data Fig. 6 | SomArchon enables both local dendritic and population imaging of neurons in multiple brain regions in vivo.** **a**, Fluorescence images of selected FOV in motor cortex (left) with selected ROIs corresponding to somas of 3 neurons (right) (*n* = 1 FOV from 1 mouse). Scale bar, 50 µm. **b**, Representative fluorescence traces from **a** with detected spikes (black ticks). **c**, Fluorescence image of a hippocampal neuron expressing SomArchon with ROIs selected at the soma and on 4 proximal dendrites (*n* = 1 neuron from 1 mouse). Scale bar, 20 µm. **d**, Optical voltage traces from the selected ROIs shown in **c**. **e**, Fluorescence image of a striatal neuron expressing SomArchon with ROIs selected at the soma and on 3 proximal dendrites (*n* = 1 neuron from 1 mouse). Scale bar, 20 µm. **f**, Optical voltage traces from the selected ROIs shown in **e**. Black arrows in **d**, **f** highlight instances in which dendritic voltages differed visibly from those on the soma. **g**–**k**, In vivo population voltage imaging in the hippocampus CA1 region (*n* = 14 FOVs from 3 mice). **g**, **i**, Average intensity projection image for each video (top), with identified ROIs (bottom). **h**, **j**, Optical voltage traces for each neuron shown in **g**, **i**, respectively, with colours matching corresponding ROI colours. Panels show 1.2 s of simultaneously recorded voltage for all neurons (left), and a period with prominent spikes (right). Image acquisition rate for all recordings, 826 Hz.



| | Firing Rate (Hz) | Area (um²) | ISI (mean+/- SD) |
|---|---|---|---|
| **Cell 1** | 2 | 187.59 | 403.8+/-881.5 |
| **Cell 2** | 1.6686 | 144.073 | 530.2+/-953.4 |
| **Cell 3** | 2.7771 | 172.803 | 300.4+/-542.4 |
| **Cell 4** | 2.9167 | 175.338 | 273.3+/-509.6 |
| **Cell 5** | 7.933 | 175.76 | 124.4+/-198.3 |
| **Cell 6** | 13.2778 | 89.993 | 71.3+/-88.6 |
| **Cell 7** | 15.1389 | 69.29 | 64.8+/-87.6 |
| **Cell 8** | 11.7222 | 86.612 | 83.7+/-98.9 |
| **Cell 9** | 29.95 | 131.82 | 33.2+/-39.8 |
| **Cell 10** | 2.0167 | 117.455 | 462.6+/-586.2 |
| **Cell 11** | 2.7667 | 149.143 | 351.2+/-470.7 |
| **Cell 12** | 5.4271 | 361.66 | 182.4+/-133.1 |
| **Cell 13** | 14.3626 | 181.253 | 69.1+/-96.8 |
| **Cell 14** | 23.1753 | 191.393 | 43.0+/-44.0 |

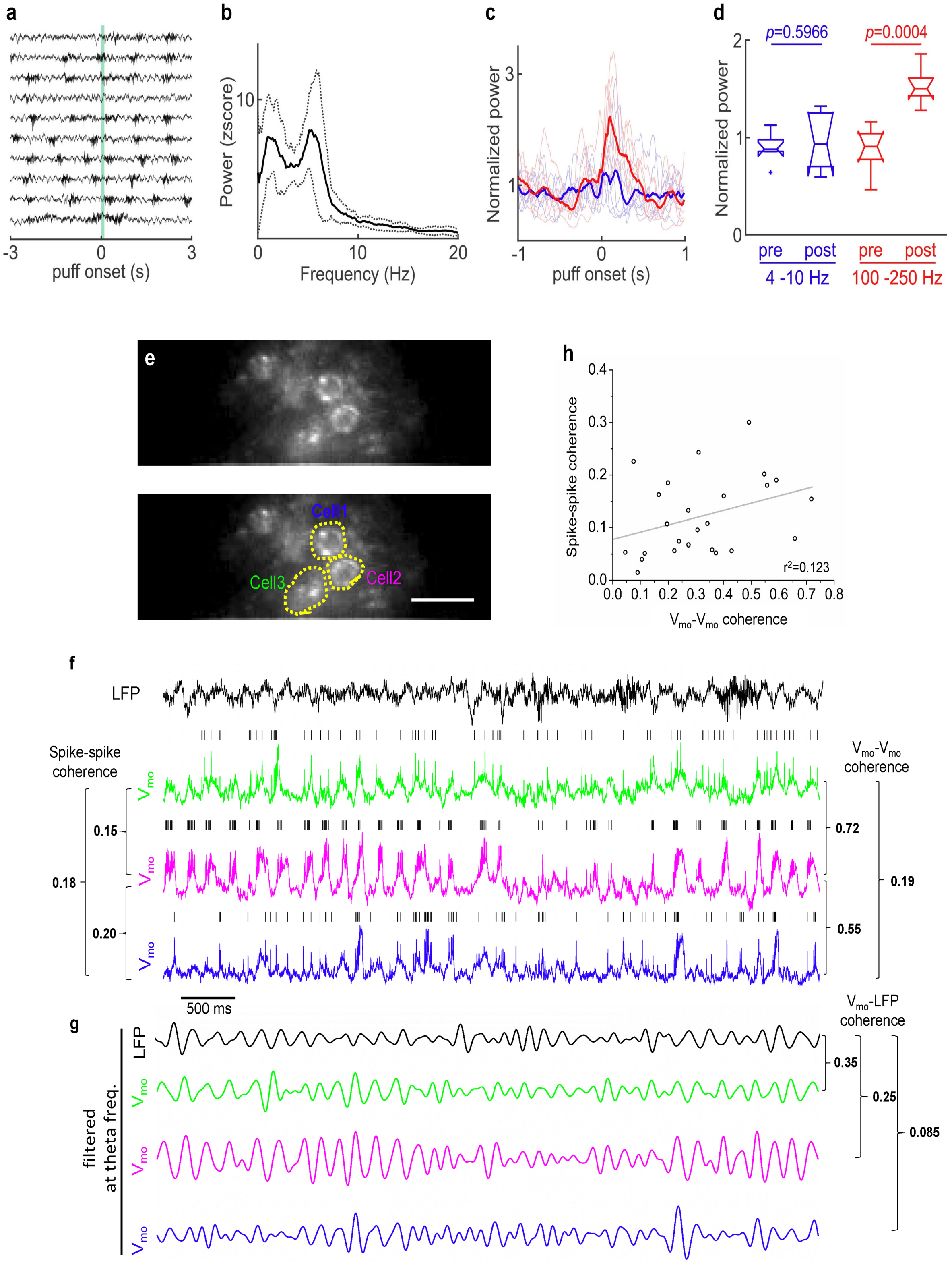
**Extended Data Fig. 7** | See next page for caption.

**Extended Data Fig. 7 | Properties of striatal neurons and movement thresholds.** **a**, Average firing rate, size, and interspike interval (ISI) for 14 neurons recorded in 9 FOV in 2 mice. Cells simultaneously recorded in the same FOV are colour-coded (blue, red and green). Cells in rows with a white background were recorded individually. **b**, Selected trace from cell 9 exhibiting spike bursting (top), and a zoomed-in view of the boxed region (bottom). A.U., arbitrary unit. Identified spikes are indicated by the marks on top of the trace. **c**, Single frame images for FOVs with multiple neurons, colour-coded as in **a**. Scale bars, 20 μm. **d**, Representative optical traces from two Cre-dependent SomArchon-expressing striatal cholinergic interneurons in a ChAT-Cre mouse (left; scale bar, 20 μm), recorded in 3 sessions, while mouse was awake with head fixed and navigating a spherical treadmill ($n = 2$ neurons from 1 mouse). Top trace corresponds to top neuron on left; two bottom traces correspond to bottom neuron. Image acquisition rate, 826 Hz. **e**, Histogram of instantaneous movement speeds for all FOVs shown in Fig. 3 (nine FOVs in two mice). Instantaneous movement speed was calculated as average speed during each 0.5-s time interval. Red line, threshold for low movement speed identification; green line, threshold for high movement speed identification. **f**, Histogram of instantaneous movement speed for individual FOVs analysed.

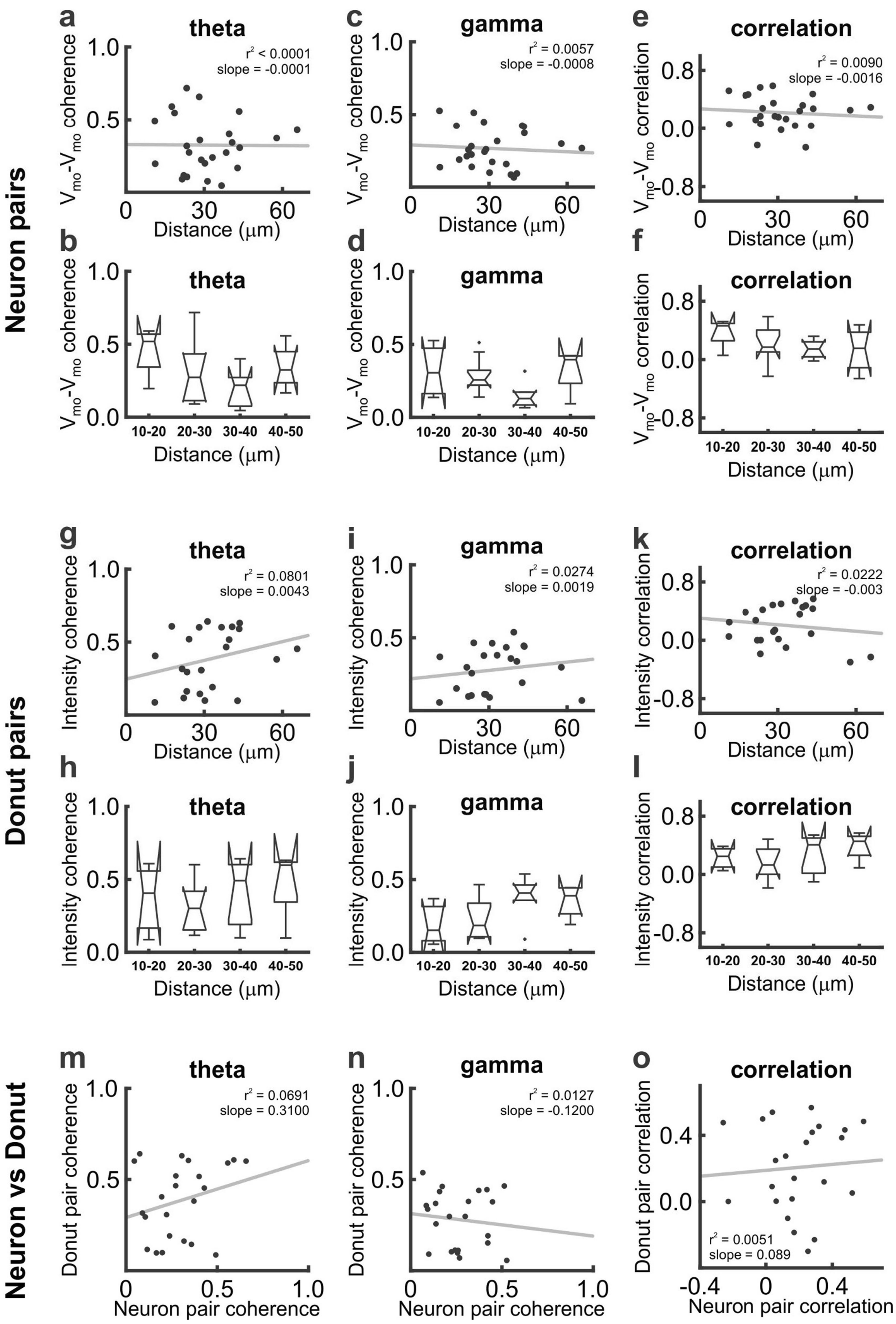


**Extended Data Fig. 8** | See next page for caption.

**Extended Data Fig. 8 | In vivo SomArchon performance over time in the striatum and hippocampus of awake mice. a**–**h**, Average fluorescence intensity, SNR per spike, and firing rates of neurons in the striatum and hippocampus of awake mice, over multiple trials. **a**–**c**, In each striatal recording session, we performed 5 trials, each 12-s long, with inter-trial intervals of 30–60 s. Average fluorescence intensity (**a**) decreased slightly; spike SNR (**b**) and firing rates (**c**) remained constant throughout the recording session (repeated-measures analysis of variance (ANOVA), $n = 6$ neurons in 5 FOVs from 1 mouse). **d**–**h**, In each hippocampal recording session, we performed 10 trials, each 6-s long, with inter-trial intervals of 20–30 s. Average fluorescence intensity (**d**) showed a slight but significant decrease across trials. SNR (**e**) decreased between the first and second trials but not afterwards, and firing rate (**f**) remained constant. Spike amplitude (**g**) fluctuated randomly over trials, and there was a significant increase in baseline noise (**h**) between the first and second trials (repeated-measures ANOVA; $^{*}P < 0.05$, post-hoc test: Tukey's HSD test, $n = 16$ neurons in 7 FOVs from 4 mice, Supplementary Table 2). Measurements were normalized to the first trial for each neuron. Box plots as in Fig. 4. **i**–**m**, A representative continuous optical trace of a hippocampal neuron over 80 s in an awake, head-fixed mouse (**i**), with zoomed-in views (**j**–**m**) at the beginning and end of the recording highlighting comparable firing rates and SNRs ($n = 16$ neurons in 7 FOVs from 4 mice).



**Extended Data Fig. 9 | Analysis of LFP and subthreshold membrane voltage oscillation in the hippocampus. a**, Example hippocampal LFP recordings from a session with ten trials, aligned to the onset of an air puff (green shading) directed to one eye in awake, head-fixed mice. **b**, LFP power spectrum shows strong theta oscillations. Mean ± s.d., $n = 10$ trials in 1 session. **c**, Oscillation power at high frequencies (100–250 Hz, red) and at theta frequencies (blue), aligned to puff onset. Each thin line represents an individual recording session, and the thick lines denote means ($n = 7$ sessions in 4 mice). **d**, Eye puff evoked a significant increase in LFP power at high frequency, but not at theta frequency (theta frequency $P = 0.5966$; high frequency $P = 0.0004$; two-tailed paired Student's $t$-test, $n = 7$ sessions in 4 mice). Box plots as in Fig. 4. **e**, Fluorescence image of a representative FOV (top) with selected ROIs (bottom). **f**, Membrane voltage recorded optically ($V_{mo}$) from neurons identified in **e**, and simultaneously recorded LFPs. Black vertical ticks above $V_{mo}$s denote spikes. Spike–spike coherence values between neurons are shown on the left and $V_{mo}$–$V_{mo}$ theta coherence values are shown on the right. **g**, Theta-frequency-filtered LFPs and $V_{mo}$s for the four traces shown in **f**. $V_{mo}$–LFP coherence values are shown on the right. **h**, Scatter plot of $V_{mo}$–$V_{mo}$ theta frequency coherence and spike–spike coherence from all neuron pairs, fitted with a linear regression ($n = 25$ pairs, $P = 0.08$, $t$-statistic, $r^2 = 0.12$).



**Extended Data Fig. 10 |** See next page for caption.

**Extended Data Fig. 10 | Pairwise coherence and correlation measures over spatial distance.** To investigate the potential of background fluorescence signals under wide-field imaging to produce shared crosstalk signals on neuron pairs, we examined the relationship of various coherence and correlation measures between neurons and background fluorescence over spatial distance. **a**, **b**, Pairwise coherence at theta frequencies between neurons. $V_{\mathrm{mo}}$–$V_{\mathrm{mo}}$ coherence did not decrease significantly with spatial distance. (**a**, $n = 25$ pairs analysed with spatial distance of 11–66 µm, centre to centre; **b**, $n = 23$ pairs within 50 µm of each other; $F = 1.44$, $P = 0.26$, one-way ANOVA). **c**, **d**, Pairwise $V_{\mathrm{mo}}$–$V_{\mathrm{mo}}$ coherence at gamma frequencies (30–50 Hz) was not dependent on spatial distance (**c**, $n = 25$ pairs; **d**, $n = 23$ pairs within 50 µm of each other; $F = 2.10$, $P = 0.13$, one-way ANOVA). **e**, **f**, Pairwise correlation between neurons did not decrease significantly with spatial distance (**e**, $n = 25$ pairs; **f**, $n = 23$ pairs within 50 µm of each other; $F = 1.00$, $P = 0.42$, one-way ANOVA). **g**–**l**, Same analysis as in **a**–**f** performed in background doughnut ROIs surrounding each neuron (Methods). Similar to results from neuron pairs, we found that theta frequency coherence between background doughnut ROIs was not dependent on spatial distance (**g**, $n = 23$ pairs; **h**, $n = 21$ pairs; $F = 0.65$, $P = 0.59$, one-way ANOVA), nor was gamma frequency coherence (**i**, $n = 23$ pairs; **j**, $n = 21$ pairs; $F = 1.93$, $P = 0.16$, one-way ANOVA), or the correlation coefficient (**k**, $n = 23$ pairs; **l**, $n = 21$ pairs; $F = 1.02$, $P = 0.41$, one-way ANOVA). **m**–**o**, The coherence between neurons and their corresponding doughnuts was not correlated at theta frequency (**m**), at gamma frequency (**n**), or for the Pearson correlation coefficients (**o**). Box plots are as in Fig. 4.

nature research

Corresponding author(s): Xue Han, Edward S. Boyden

Last updated by author(s): Apr 25, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Data were recorded using NIS-Elements Advance Research software v4.60.00, HC Image 4.5, OmniPlex system 1.6.0, and MATLAB 2014b. |
| Data analysis | Data were analyzed offline using NIS-Elements Advance Research software v4.60.00, OriginPro 8 (OriginLab), C, Excel 2016 (Microsoft), ImageJ(Fiji) 1.52i, Python 3.6.8. and 3.7.1, BoxPlotR, and MATLAB 2017a, 2018a, 2018b. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Code availability. Computer codes used to generate results for this study are available at https://github.com/HanLabBU/somarchon-imaging.
Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request; raw data essential to the work is avalible online at nature.com. Sequences of the reported proteins are available at Genbank at the following accession codes: SomArchon MN091368; SomArchon-P2A-CoChR-KV2.1motif, MN091369.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical methods were used to estimate sample size for animal studies throughout. We did not perform a power analysis, since our goal was to create a new technology; in the reference (Dell, R. B., Holleran, S. & Ramakrishnan, R. Sample size determination. ILAR. J. 43, 207–213 (2002)), as recommended by the NIH, "In experiments based on the success or failure of a desired goal, the number of animals required is difficult to estimate..." As noted in the aforementioned paper, "The number of animals required is usually estimated by experience instead of by any formal statistical calculation, although the procedures will be terminated [when the goal is achieved]." These numbers reflect our past experience in developing neurotechnologies.

Data exclusions

Voltage imaging datasets with significant motion or where no spikes were detected were excluded from analysis. Significant motion was defined as a shift for more than 20 µm in any direction. In Extended Data Figure 1i,j,k data points that corresponded to overlapping neurites were excluded. Data exclusion criteria were not pre-established.

Replication

All attempts at replication were successful. The detailed experimental protocols are provided to facilitate replication by others.

Randomization

Blinding

There were no treatment conditions to compare in this study. All recording sessions were randomly performed with different voltage sensors or in different brain regions. On recording days, cultured cells or brain slices expressing specific sensors were known. On in vivo recording days, mouse conditions were known. Voltage trace extraction and subsequent analysis were performed with the investigators unaware of specific mouse conditions. For analysis of movement modulation of striatal neuron spiking, a computer algorithm was used to identify periods with different movement parameters. For analysis of spike-phase relationships, or subthreshold membrane voltage relationships, a computer algorithm was used across all conditions. For histology,- sections were selected and images were taken from slides by a researcher not aware of the conditions or antibody used. Cells were also counted and quantified from these sections by a researcher blinded to the experimental conditions or antibody used.

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | Antibodies |
| ☒ | ☐ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology |
| ☐ | ☒ | Animals and other organisms |
| ☒ | ☐ | Human research participants |
| ☒ | ☐ | Clinical data |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | ChIP-seq |
| ☒ | ☐ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

## Antibodies

Antibodies used

Primary antibodies anti-GFAP (1:250, Clone N206/A8, Neuromab) and anti-IBa1 (1:500, 019-19741, Wako Chemicals) and secondary anti-bodies Alexa Fluor 568 (1:1000, Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, A11004, InVitrogen) 633 secondary antibodies (1:1000, Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, A21070, InVitrogen)were used in the study.

Validation

For anti-GFAP please see: http://neuromab.ucdavis.edu/datasheet/N206A_8.pdf
For anti IBa1, please see: http://www.e-reagent.com/uh/Shs.do?now=1550070200673
Validation methods are references in Methods.

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Species, strain, sex, and age are reported for each experiment in the Methods

Wild animals

The study did not involve wild animals

| | |
|---|---|
| Field-collected samples | This study did not involve samples collected in the field. |
| Ethics oversight | Boston University Institutional Animal Care and Use and Biosafety Committee, Massachusetts Institute of Technology Institutional Animal Care and Use and Biosafety Committee |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

LETTER

https://doi.org/10.1038/s41586-019-1654-9

# Organoid single-cell genomic atlas uncovers human-specific features of brain development

Sabina Kanton[1,7], Michael James Boyle[1,7], Zhisong He[1,2,7]*, Malgorzata Santel[1], Anne Weigert[1], Fátima Sanchís-Calleja[1,2], Patricia Guijarro[3], Leila Sidow[1], Jonas Simon Fleck[2], Dingding Han[3], Zhengzong Qian[3], Michael Heide[4], Wieland B. Huttner[4], Philipp Khaitovich[1,3,5], Svante Pääbo[1], Barbara Treutlein[1,2]* & J. Gray Camp[1,6]*

**The human brain has undergone substantial change since humans diverged from chimpanzees and the other great apes[1,2]. However, the genetic and developmental programs that underlie this divergence are not fully understood. Here we have analysed stem cell-derived cerebral organoids using single-cell transcriptomics and accessible chromatin profiling to investigate gene-regulatory changes that are specific to humans. We first analysed cell composition and reconstructed differentiation trajectories over the entire course of human cerebral organoid development from pluripotency, through neuroectoderm and neuroepithelial stages, followed by divergence into neuronal fates within the dorsal and ventral forebrain, midbrain and hindbrain regions. Brain-region composition varied in organoids from different iPSC lines, but regional gene-expression patterns remained largely reproducible across individuals. We analysed chimpanzee and macaque cerebral organoids and found that human neuronal development occurs at a slower pace relative to the other two primates. Using pseudotemporal alignment of differentiation paths, we found that human-specific gene expression resolved to distinct cell states along progenitor-to-neuron lineages in the cortex. Chromatin accessibility was dynamic during cortex development, and we identified divergence in accessibility between human and chimpanzee that correlated with human-specific gene expression and genetic change. Finally, we mapped human-specific expression in adult prefrontal cortex using single-nucleus RNA sequencing analysis and identified developmental differences that persist into adulthood, as well as cell-state-specific changes that occur exclusively in the adult brain. Our data provide a temporal cell atlas of great ape forebrain development, and illuminate dynamic gene-regulatory features that are unique to humans.**

Bulk genomic measurements in primary brain tissue from adult humans, chimpanzees and other apes[3–6], as well as from developing rhesus macaques[7,8], have identified molecular features that appear to be specific to the human brain. These studies have largely been limited by ensemble averaging and it has been difficult to perform similar experiments in developing great apes owing to the lack of available tissue. Cerebral organoids[9] grown from great ape induced pluripotent stem cells (iPSCs)[10] offer the potential to study the evolution of human brain development in controlled culture environments. Previous single-cell sequencing studies have shown that human organoids can recapitulate many aspects of in vivo cortex development[11–14], and comparisons with other primate organoids can be used to identify human-specific gene expression[15,16]. However, the gene-regulatory mechanisms that distinguish humans from the other primates are unclear. Here we use single-cell RNA sequencing (scRNA-seq) together with accessible chromatin profiling to understand human cerebral organoid development and explore how human cortical gene-expression programs have diverged from chimpanzee. We further analysed adult prefrontal cortex tissue using single-nucleus RNA-seq (snRNA-seq) to reveal the potential and limits of cerebral organoids to study human-specific expression patterns.

We first used scRNA-seq (10x Genomics) to profile cell composition across a time course of human organoid development from pluripotency to four months using embryonic stem cells (H9) and an iPSC (409b2) line (Fig. 1a, Extended Data Fig. 1; 43,498 cells). We constructed a force-directed *k*-nearest neighbour graph[17] based on pseudocells to visualize the temporal progression (Fig. 1b). We tracked a progression through stem cell states during the first 15 days, and by 1 month, cells diversified into neural progenitor cells (NPCs) of multiple brain regions including the forebrain (dorsal and ventral telencephalon and diencephalon), midbrain (mesencephalon), hindbrain (rhombencephalon) and retina. By two months, excitatory and inhibitory neuronal fates had differentiated, and by four months, astrocytes had emerged. These observations were based on inspection of marker genes, comparisons to transcriptomes from regions and cells from the developing human brain[18,19], and analysis of spliced and unspliced transcripts[20] (Fig. 1c, d, Extended Data Fig. 2).

We next analysed the reproducibility of gene-expression patterns across pluripotent stem cell (PSC) lines from different human individuals in two-month-old organoids (Fig. 1e, Extended Data Fig. 3). We identified cells on the neuronal lineage (49,153 cells), classified distinct cell types from different brain regions, analysed differentiation trajectories and ordered ventral and dorsal telencephalic cells along pseudotimes (Fig. 1e–g, Extended Data Fig. 3). Consistent with previous studies[12,21], we found that each inducible PSC (iPSC) line contributed cells to multiple differentiation trajectories; however, the proportions of cells in each trajectory varied across organoid and line (Fig. 1f, Extended Data Fig. 3). Nevertheless, gene-expression patterns of each trajectory across the lines were highly correlated and cells representing each region clustered together (Fig. 1h, i, Extended Data Fig. 3). These data provide a temporally and pseudotemporally resolved gene-expression atlas of the earliest stages of human brain development and provide a baseline for identifying human-specific gene expression.

We next generated a gene-expression atlas of chimpanzee organoid development to dissect features that differ from that of humans (Fig. 2a, Extended Data Fig. 4; 36,884 cells). We identified dorsal and ventral telencephalon trajectories, as well as rhombencephalic cell populations in chimpanzee organoids (Fig. 2b). We used time warping to align human and chimpanzee cortical pseudotimes and observed that the latest pseudotime points in chimpanzee failed to map to a human counterpart (Fig. 2c). We found that chimpanzee neurons express higher levels of neuron maturation-related genes (Fig. 2d) and confirmed differences in maturation using scRNA-seq data from additional human and chimpanzee individuals[15,16], as well as from macaque (Extended Data Figs. 5, 6). In addition, more astrocytes were observed in chimpanzee organoids compared to humans at four months (Fig. 2e). We observed more pronounced upper- and deep-layer cortical neuron specification

---

[1]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [2]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland. [3]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai, China. [4]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [5]Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology, Moscow, Russia. [6]Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland. [7]These authors contributed equally: Sabina Kanton, Michael James Boyle, Zhisong He. *e-mail: zhisong.he@bsse.ethz.ch; barbara.treutlein@bsse.ethz.ch; grayson.camp@iob.ch

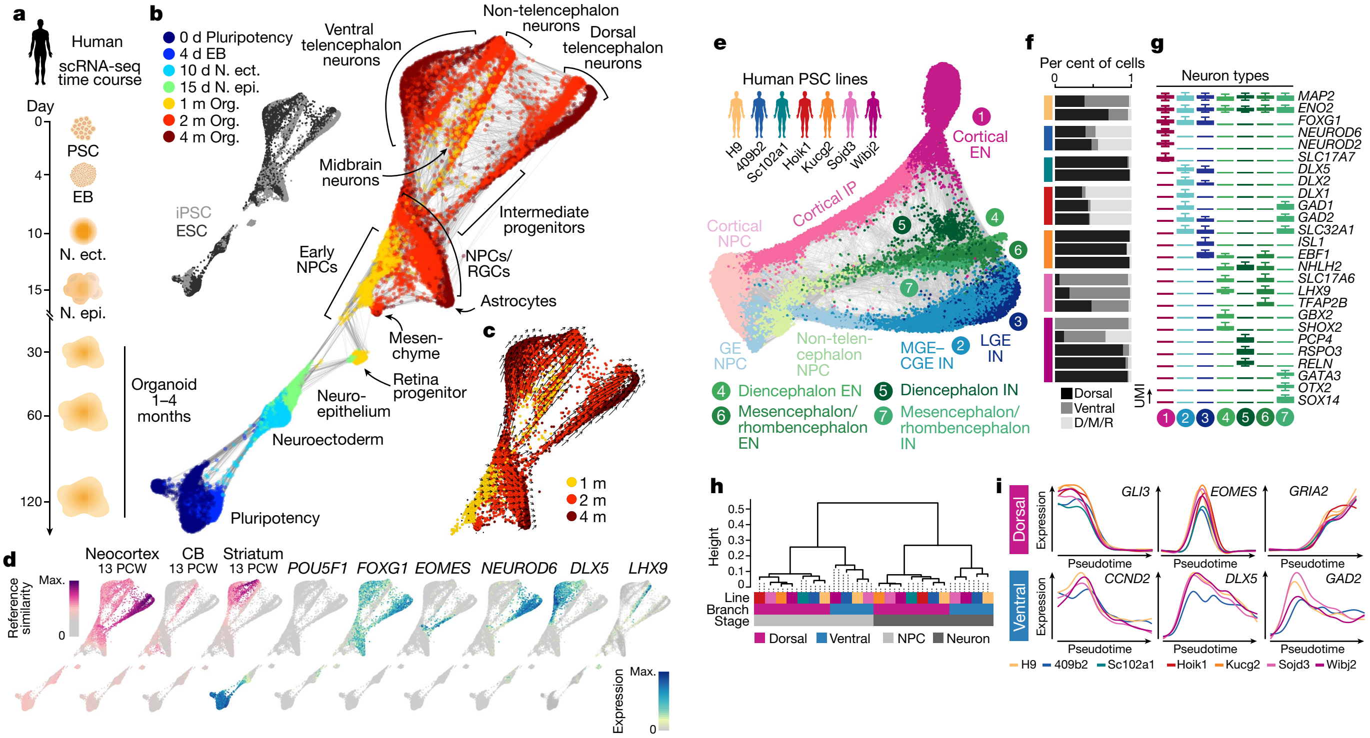


**Fig. 1 | Reconstructing human cerebral organoid differentiation from pluripotency. a**, scRNA-seq was performed on iPSC- and embryonic stem cell (ESC)-derived cells (43,498) at different time points during cerebral organoid differentiation from pluripotency. EB, embryoid body; N. ect., neuroectoderm; N. epi., neuroepithelium. **b**, All time points were combined, pseudocells (11,427) were constructed and the differentiation trajectory was reconstructed using SPRING[17]. Pseudocells are coloured by time point (main image) or cell line (top left). Org., organoid; RGC, radial glial cell. **c**, RNA velocity analysis[20] supports differentiation of NPCs into distinct regions of the developing human brain. **d**, Left, SPRING plot coloured (magenta) by reference similarity spectrum (RSS) to bulk RNA-seq data generated from diverse brain regions at different time points (Allen Brain Atlas)[18]. Shown are the tissues and time points with maximum correlation. CB, cerebellum. Right, SPRING plot coloured (cyan) by marker gene expression. PCW, post-conception weeks. **e**, SPRING reconstruction based on the RSS of organoid scRNA-seq data from six iPSC lines and one ESC (H9) line (49,153 cells), with clusters coloured by cell type. CGE, caudal ganglionic eminence; MGE, medial ganglionic eminence; LGE, lateral ganglionic eminence; IN, inhibitory neuron; EN, excitatory neuron; IP, intermediate progenitors; GE, ganglionic eminence. **f**, Proportion of cells per organoid that are in the dorsal or ventral telencephalon or diencephalon, mesencephalon and rhombencephalon (D/M/R) neuronal branches. **g**, Box plots (outliers removed) showing relative expression of marker genes for major neuron populations that emerge in the human cerebral organoids. Boxes represent interquartile range (IQR) and whiskers represent minimum and maximum with outliers removed. UMI, unique molecular identifier. **h**, Dendrogram based on pairwise correlations between cells from different lines, branches or stages based on pseudotime-dependent gene-expression patterns. **i**, Pseudotemporal expression patterns of neuronal differentiation markers for the dorsal (top) and ventral (bottom) telencephalon trajectories for each line.

in chimpanzee and macaque organoids relative to human organoids at the same time point on the basis of expression, but this was not consistent across lines, organoids and protocols (Extended Data Fig. 6). Our data suggest that delayed maturation of the human brain[5,22,23] is observed in organoids and can be traced back to early stages of development, consistent with previous reports comparing human and macaque brain development in vivo[7] and in in vitro 2D cultures[24,25].

We next aimed to detect human-specific changes in gene expression in the developing cortex. We first inspected the expression of duplicated or rearranged genes (Supplementary Table 7) and detected 23 of 24 genes included in the annotation, 4 of them specific to G2M phase progenitors of the telencephalon (*ARHGAP11B*, *FAM72B*, *FAM72C* and *FAM72D*)[26]. We next aligned all human, chimpanzee and macaque reads to a consensus genome and aligned dorsal telencephalic trajectories between the species (Fig. 2f, Extended Data Fig. 7). We then searched for genes that were differentially expressed specifically on the human branch, and found that most of the observed human-specific deviations from chimpanzee and macaque were expression gains (Fig. 2g); this was also observed for chimpanzee-specific changes (Extended Data Fig. 7). We propose that this is because it is more deleterious to lose a highly conserved gene-expression pattern than it is to gain the expression of a new gene. Of the 98 identified differentially expressed genes, 96 clustered into 7 pseudotemporal patterns (Fig. 2h–j), with clusters 1, 2 and 3 specifically enriched in human radial glia, intermediate progenitors and neurons, respectively (Fig. 2h). Genes with human-specific expression gain are associated with diverse biological processes including proliferation of radial glia, neuron migration and neurite formation, and are localized to different clusters of maturing neurons (Fig. 2i, j, Extended Data Fig. 7). When comparing our results to previously published data on human and great ape organoids and fetal brains[15,16], we find strong overlap between datasets (Extended Data Fig. 7). Using cells of ventral telencephalon identity, we find 92 genes differentially expressed between human and chimp—17% of them distinct from those differentially expressed in the cortex (Extended Data Fig. 7). In sum, this analysis identifies human-specific changes in gene expression that may be specific to certain cell states within the developing human forebrain.

To identify potential regulatory mechanisms, we performed bulk and single-cell accessible chromatin profiling (scATAC-seq; Fluidigm C1) along the organoid-differentiation time course in human and chimpanzee, including that of microdissected cortical regions (Fig. 3a–c, Extended Data Fig. 8). We analysed cell heterogeneity, ordered cells in pseudotime and dynamically monitored transcription factor binding motifs and chromatin accessibility over the differentiation path (Fig. 3d, Extended Data Fig. 8). We searched for differential accessibility between human and chimpanzee cortical NPCs and neurons, and found that 7% and 9% of accessible peaks showed increased and decreased accessibility in humans, respectively (Fig. 3e), with many differentially accessible peaks being specific to either NPCs (53.8%) or neurons (33.8%). Differential accessibility peaks are enriched for

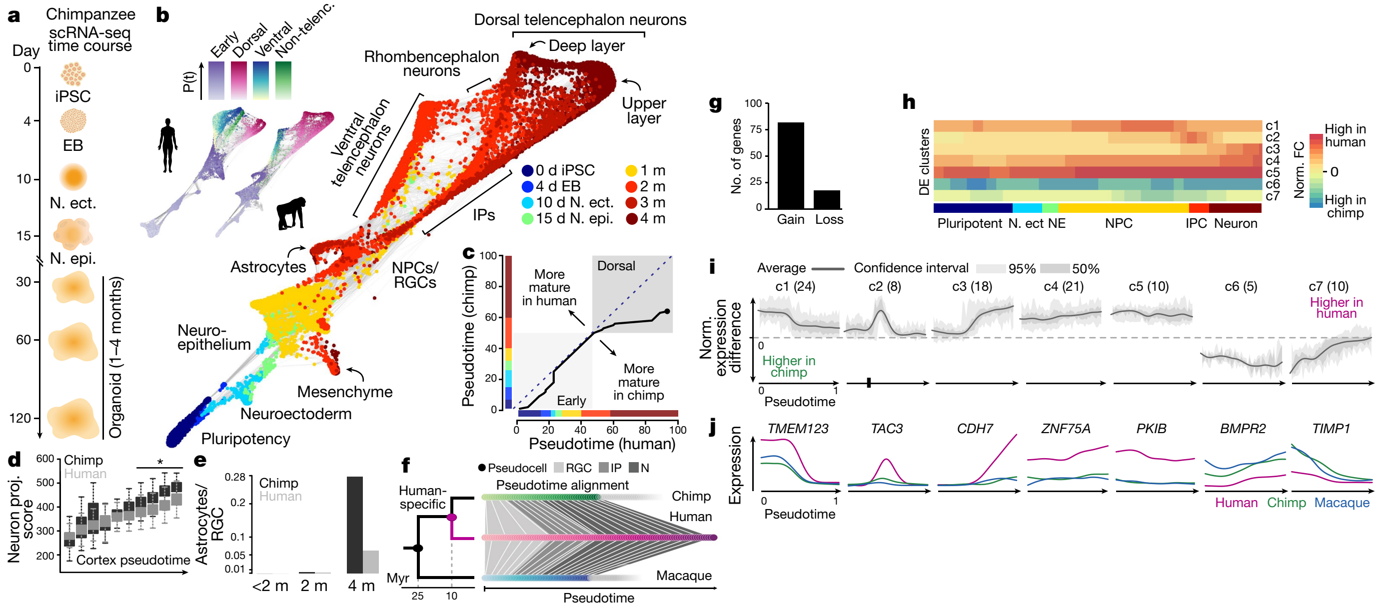


**Fig. 2 | Comparing human and chimpanzee organoids reveals human-specific features of cortex development. a**, scRNA-seq was performed on chimpanzee iPSC-derived cells at different time points during cerebral organoid differentiation from pluripotency. **b**, The differentiation trajectory was reconstructed using SPRING. Pseudocells (9,647) are coloured by time point (main image) or trajectory pseudotimes (P(t)) (top left, alongside human data). Non-telenc., non-telencephalon. **c**, Alignment of human and chimpanzee pseudotimes after combining pseudocells from the early stages and the dorsal forebrain lineage. **d**, Box plots (outliers removed) showing cumulative expression of neuron projection-related genes in human and chimpanzee along unaligned cortical pseudotimes. Boxes represent IQR and whiskers represent minimum and maximum with outliers removed. **e**, Astrocytes identified by scRNA-seq in organoids at different time points, normalized by the corresponding number of radial glia. **f**, Schematic of pseudotime alignment from cortical NPC to deeper layer neurons in human, chimpanzee and macaque. N, neuron. **g**, Number of differentially expressed genes in human versus chimpanzee and macaque comparison grouped by gain or loss of expression in humans. **h**, Average human–chimpanzee differential expression (DE) patterns along the trajectory from pluripotent cells to cortical neurons shown for the seven clusters of differentially expressed genes. NE, neuroepithelium; IPC, intermediate progenitor cell; Norm. FC, normalized fold-expression change. **i**, **j**, Distinct human-specific pseudotemporal expression patterns (**i**, cluster number with number of genes in parenthesis) and exemplary genes (**j**) for each of the seven clusters.

distinct biological processes (Fig. 3f), and many differentially accessible peaks are accessible specifically in organoids (Fig. 3g) and drive reporter expression in the mouse developing forebrain[27] (Extended Data Fig. 9). Most differentially accessible peaks were located in non-protein-coding regions distal to the promoter (Fig. 3h). The majority of genes that were differentially expressed between human and chimpanzee along the dorsal telencephalon trajectory have one or more differentially accessible peaks nearby (Fig. 3h), and are more likely to be near a differentially accessible region than non-differentially expressed genes (Extended Data Fig. 9). Differentially accessible peaks are also significantly enriched for single nucleotide changes (SNCs) that are fixed in humans and distinct from other primates (Fig. 3i), some of which are predicted to generate new or disrupt existing binding sites for organoid-expressed transcription factors (Extended Data Fig. 9). We annotated differentially accessible peaks and identified organoid-specific potential regulatory regions near differentially expressed genes that have human-fixed SNCs, have undergone accelerated evolution in humans (human accelerated regions (HARs)) or overlap conserved regions that have been deleted in humans (hCONDELs) (Fig. 3j, Supplementary Table 11). This analysis identified many regions specific to NPCs and neurons (Extended Data Fig. 9), and we found 62 HARs that overlap differentially accessible peaks, one of which is near cadherin 7 (*CDH7*), a gene with human-specific expression in neurons (Fig. 3k).

Finally, we wanted to know whether the human-specific gene-expression patterns observed in the developing cortex persist into adulthood. We generated snRNA-seq data from post-mortem prefrontal cortex tissues of three humans, two chimpanzees, one bonobo and three macaques by isolating nuclei from sequential sections from basal to apical positions[3] (Fig. 4a). Nuclei clustered into different populations of neurons, astrocytes, oligodendrocytes, microglia and endothelial cells (Fig. 4a, b, Extended Data Fig. 10). Genes specifically expressed in neurons showed higher sequence conservation than genes specific to other cell types (Fig. 4c), astrocytes had the largest number of human-specific differentially expressed genes (Fig. 4d) and some genes with human-specific differential expression in excitatory neurons showed layer specificity (Fig. 4e). A substantial fraction of the genes expressed in excitatory or inhibitory neurons in the adult cortex were also detected in their counterparts in the organoid telencephalon (Fig. 4f, Extended Data Fig. 10). Notably, relatively few genes were commonly detected as differentially expressed (human versus chimp) in organoid and adult excitatory (53 of 354) or inhibitory (13 of 217) neurons (Fig. 4g), and differentially expressed genes detected ubiquitously in the organoid cortex show stronger consistency with adult neurons compared with organoid NPC- or neuron-specific genes (Extended Data Fig. 10). Examples of human-specific differentially expressed genes in adult cell classes (Fig. 4h) include differentially expressed genes detected in developing neurons, such as *COL6A1*, which has been shown to have a protective role in ageing neurons[28]. These results suggest that, with some exceptions, transcriptome differences between human and chimpanzee cortical cells are linked to developmental stages.

In sum, our data illuminate expression and regulatory features of cell states that are uniquely human, and provide an extensive resource to guide exploration into the gene-regulatory mechanisms that distinguish the developing human and chimpanzee brains—some of which persist into adulthood.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1654-9.

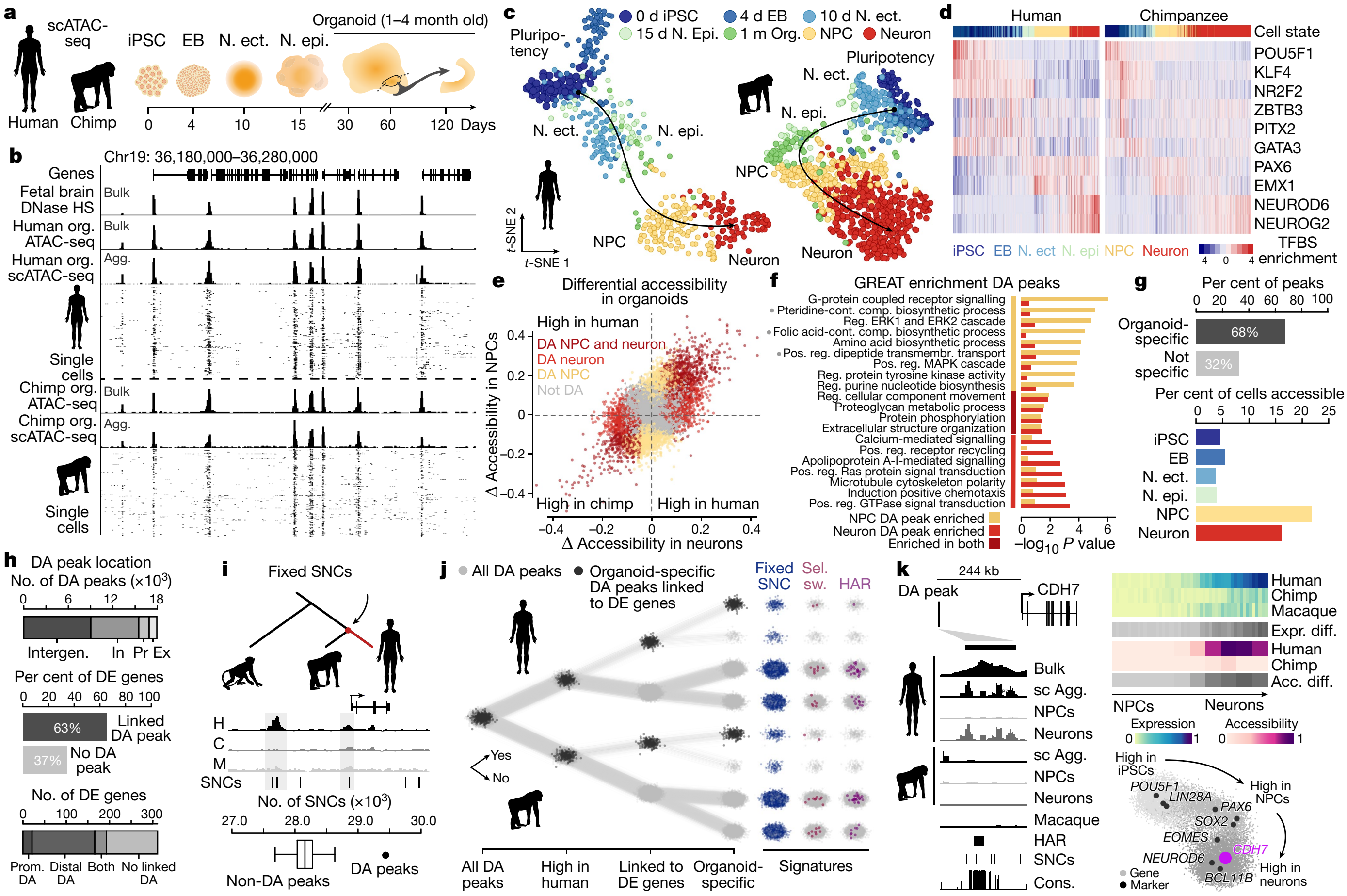


**Fig. 3 | scATAC-seq reveals chromatin accessibility dynamics during cortex development and evolution. a**, scATAC-seq was performed at different time points of human and chimpanzee cerebral organoid development from pluripotency to four-month-old organoids (microdissected cortical regions). **b**, Bulk, single-cell and aggregated (Agg.) scATAC-seq profiles from two-to-four-month-old organoids at a representative locus. **c**, ChromVAR[29]-based *t*-stochastic neighbour embedding (*t*-SNE) of human (left, 518 cells) and chimpanzee (right, 908 cells) with cells coloured by time point or cell state. **d**, Heat maps showing binding-motif enrichment for selected transcription factors in cells ordered in pseudotime. **e**, Differentially accessible (DA) peaks detected between human and chimpanzee NPCs and neurons. **f**, Enrichment of selected biological-process gene ontology (GO) terms associated with differentially accessible peaks in NPCs (gold) or neurons (light red) relative to all accessible organoid peaks. Grey dots indicate significantly enriched terms (hypergeometric test, FDR <0.05 and twofold region-based enrichment). Cont., containing; comp., compound; reg., regulation; pos., positive. **g**, Top, percentage of differentially accessible peaks that are accessible only in organoids (organoid-specific) or also at earlier stages (not specific). Bottom, percentage of human cells from each cell state accessible at differentially accessible peaks. **h**, A differentially accessible peak was linked to the nearest expressed gene within 1 Mb of that gene's transcription start site. Top, number of differentially accessible peaks located in exonic (Ex), promoter (Pr), intronic (In) or intergenic (Intergen) regions. Middle, percentage of differentially expressed genes linked with differentially accessible peaks. Bottom, proportion of differentially expressed genes with a differentially accessible peak at the promoter region (Prom. DA), distal region (Distal DA), both, or no linked differentially accessible peak. **i**, Numbers of SNCs derived and fixed in humans overlapping differentially accessible peaks and non-differentially accessible peaks (randomly sampled 2,000 times to match the number and average accessibility of differentially accessible peaks). DE, differentially expressed. **j**, Differentially accessible peak annotation noting overlap with selective sweeps (sel. sw.) or HARs. **k**, A differentially accessible peak close to the differentially expressed gene *CDH7* overlaps fixed SNCs and a HAR. Bottom right, a gene correlation network with *CDH7* highlighted. Expr., expression; Acc., accessibility; sc Agg., single-cell aggregated; Cons., conservation.

Received: 28 November 2018; Accepted: 6 September 2019;

1. Sousa, A. M. M., Meyer, K. A., Santpere, G., Gulden, F. O. & Sestan, N. Evolution of the human nervous system function, structure, and development. *Cell* **170**, 226–247 (2017).
2. Martin, R. D. Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* **293**, 57–60 (1981).
3. He, Z. et al. Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. *Nat. Neurosci.* **20**, 886–895 (2017).
4. Sousa, A. M. M. et al. Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017).
5. Somel, M. et al. Transcriptional neoteny in the human brain. *Proc. Natl Acad. Sci. USA* **106**, 5743–5748 (2009).
6. Konopka, G. et al. Human-specific transcriptional networks in the brain. *Neuron* **75**, 601–617 (2012).
7. Zhu, Y. et al. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018).
8. Reilly, S. K. et al. Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
9. Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
10. Marchetto, M. C. N. et al. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**, 525–529 (2013).
11. Camp, J. G. et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl Acad. Sci. USA* **112**, 15672–15677 (2015).
12. Quadrato, G. et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).
13. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
14. Amiri, A. et al. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* **362**, eaat6720 (2018).
15. Mora-Bermúdez, F. et al. Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife* **5**, e18683 (2016).

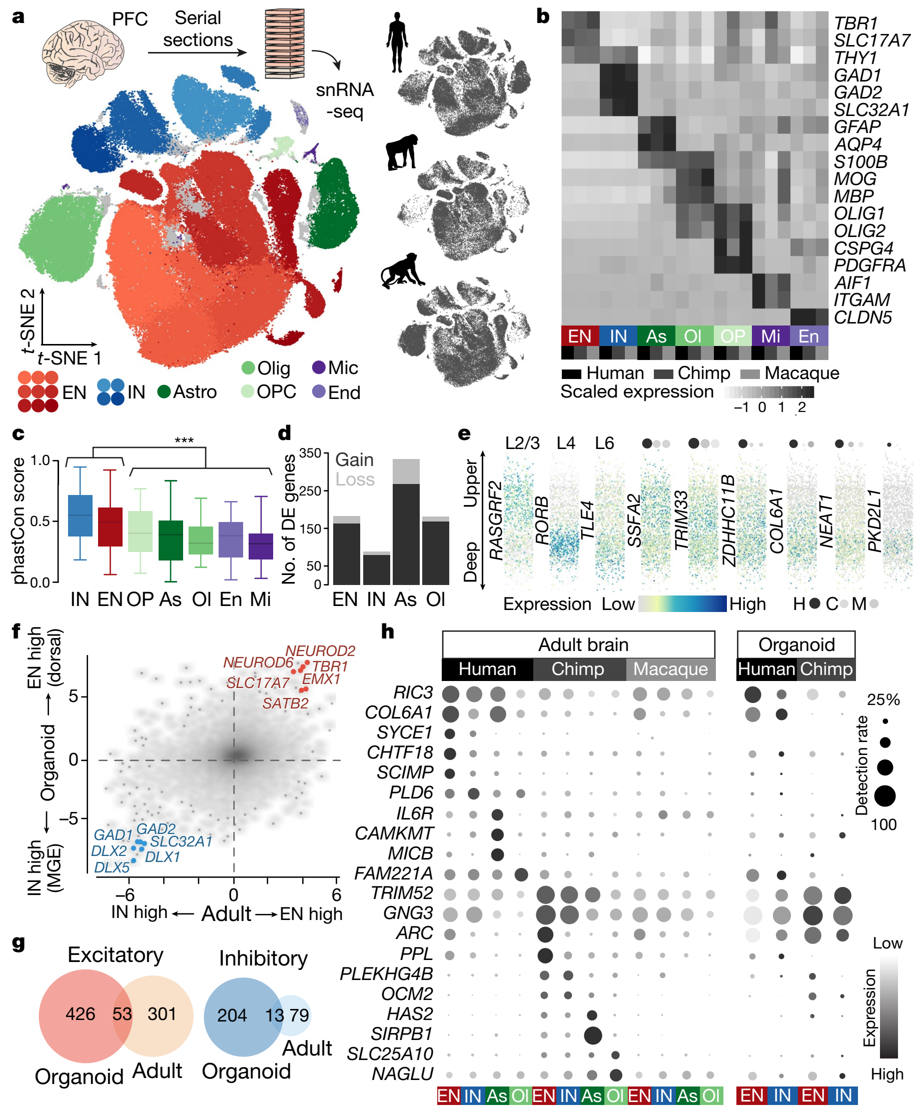


**Fig. 4 | Adult cortex snRNA-seq reveals shared and distinct patterns of gene-expression change compared to organoids. a**, snRNA-seq was performed on sliced tissue cubes dissected from adult human, chimpanzee or bonobo, and macaque frozen prefrontal cortex tissue. Projection of integrated data shows different clusters of major cell classes in the different species (50,035 human, 33,847 chimpanzee or bonobo, and 50,403 macaque nuclei). Astro, astrocyte; Olig, oligodendrocyte; OPC, oligodendrocyte precursor cell; Mic, microglia; End, endothelial cell. **b**, Average cluster expression separated by species reveals similar patterns of marker gene expression for seven cell classes. As, astrocyte; Ol, oligodendrocyte; OP, oligodendrocyte precursor; Mi, microglia; En, endothelial cell. **c**, Genomic conservation (average phastCon scores) of markers for seven cell classes (***$P < 0.0001$, two-sided Wilcoxon's rank-sum test). Boxes represent IQR and whiskers represent 1.5 × IQR. **d**, Number of human-specific differentially expressed genes in each of the four major cell classes. **e**, Predicted laminar expression patterns of three layer (L) markers (left) and six human-specific differentially expressed genes (right) in adult excitatory neurons with size and shade of dots above the expression pattern plot showing detection rates and average expression levels, respectively, in the three species. **f**, $\log_2$(fold change) between excitatory and inhibitory neurons shows consistency between organoid dorsal and ventral forebrain and adult tissue. **g**, Overlap of human–chimpanzee differentially expressed genes detected in adult neurons and organoid trajectories. **h**, Dot plot showing expression patterns and detection rates across adult and organoid cell classes for human-specific differentially expressed genes in adult cell classes.

16. Pollen, A. A. et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell* **176**, 743–756 (2019).
17. Weinreb, C., Wolock, S. & Klein, A. M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 (2018).
18. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
19. Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
20. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
21. Renner, M. et al. Self-organized developmental patterning and differentiation in cerebral organoids. *EMBO J.* **36**, 1316–1329 (2017).
22. Bakken, T. E. et al. A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367–375 (2016).
23. Leigh, S. R. Brain growth, life history, and cognition in primate and human evolution. *Am. J. Primatol.* **62**, 139–164 (2004).
24. Marchetto, M. C. et al. Species-specific maturation profiles of human, chimpanzee and bonobo neural cells. *eLife* **8**, e37527 (2019).
25. Otani, T., Marchetto, M. C., Gage, F. H., Simons, B. D. & Livesey, F. J. 2D and 3D stem cell models of primate cortical development identify species-specific differences in progenitor behavior contributing to brain size. *Cell Stem Cell* **18**, 467–480 (2016).
26. Florio, M. et al. Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *eLife* **7**, e32332 (2018).
27. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
28. Gregorio, I., Braghetta, P., Bonaldo, P. & Cescon, M. Collagen VI in healthy and diseased nervous system. *Dis. Model. Mech.* **11**, dmm032946 (2018).
29. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Pluripotent stem cell lines and organoid culture.** We acquired 6 human induced pluripotent stem cell (iPSC) lines (Sojd3, Hoik1, Kucg2, Wibj2 from the HipSci resource[30]; h409b2 from the RIKEN BRC cell bank[11]; Sc102a1 from System Biosciences), one human ES cell line (H9, WiCell)[31], three chimpanzee iPSC lines (SandraA[15]; PR818-5[15], originally generated by the Gage laboratory and kindly provided to us by the R. Livesey group; JoC, generated in this study), one bonobo iPSC line (Bokela, generated in this study), one ES macaque cell line (MN1[25], kindly provided through the R. Livesey group from E. Curnow) and one orangutan cell line (Toba[15]). The iPSC line JoC (chimpanzee, Tchimpounga Sanctuary) was reprogrammed from blood cells (primary lymphocytes) using plasmid-based reprogramming[32] and Bokela (bonobo, Zoo Leipzig) was reprogrammed from fibroblasts using the StemMACS mRNA transfection kit (Miltenyi Biotec). Cell lines were validated for pluripotency markers by immunohistochemical stainings using the Human Pluripotent Stem Cell 3-Colour Immunohistochemistry Kit (R&D Systems, SC021) and were differentiated into the three different germ layers using the Human Pluripotent Stem Cell Functional Identification kit (R&D Systems) and StemMACS Trilineage Differentiation Kit (Miltenyi Biotec). Karyotyping was carried out using Giemsa banding at the Stem Cell Engineering facility, a core facility of CMCB at Technische Universität Dresden, and karyotypes were found to be normal. Cell lines were cultivated using standard feeder-free conditions in mTeSR1 (StemCell Technologies) or StemMACS iPS-Brew XF (Myltenyi Biotec) on matrigel-coated plates and differentiated into cerebral organoids using a whole organoid differentiation protocol[33] (Supplementary Table 1). Toba orangutan iPSCs were maintained on feeder cells and cultivated in ReproMix (1:1 mix Primate ES medium (Reprocell) and TeSR2 (StemCell Technologies), 8 μg/ml bFGF (Sigma), 50 U/ml penicillin/streptomycin (Thermo Fisher)). Colonies were picked from feeder cells to collect cells for generating EBs. Cell lines were regularly tested for mycoplasma using PCR validation (Venor GeM Classic, Minerva Biolabs) and found to be negative. Permission to work with human and non-human primate iPSC lines was obtained through the Sächsisches Staatsministerium fur Umwelt und Landwirtschaft (Az.: 55- 8811.72/26, Az.: 55-8811.72/26/382, Az.: 55-8811.72/26/393, 54-8452/26/7). The use of human ESCs for the generation of cerebral organoids was approved by the ethics committee of the Robert Koch Institut (https://www.rki.de/DE/Content/Gesund/Stammzellen/Register/reg-20161027-Paeaebo.html) as well as by the Ethics committee of northwest and central Switzerland (2019-01016) and the Swiss federal office of public health.

**Single-cell RNA-seq data generation.** A summary of all single-cell experiments can be found in Supplementary Table 1. For organoid experiments (1 month, 2 months, 3 months, 4 months), whole organoids were dissociated for generating single-cell gene-expression libraries. In brief, organoids were transferred to HBSS (without $Ca^{2+}$ and $Mg^{2+}$, −/−) and cut into two pieces to clear away debris from the centre of the organoid (2–3 washes in total). Organoid pieces were then dissociated using Neural dissociation kit (P) using Papain-based dissociation (Miltenyi Biotec). Organoid pieces were incubated in Papain at 37 °C (enzyme mix 1) for an initial 15 min, followed by addition of Enzyme A (enzyme mix 2) to the Papain mix. Organoid pieces were then triturated using wide-bore 1,000-ml tips and incubated for additional intervals of 5–10 min with triturations between the incubation steps, amounting to a total Papain incubation time of approximately 45 min. Cells were filtered through a 30-μm strainer and washed, centrifuged for 5 min at 300*g* and washed 3 times with HBSS (−/−). Cells were then analysed using Trypan Blue assay, counted using the automated cell counter Countess (Thermo Fisher), and diluted for an appropriate concentration to obtain approximately 6,000 cells per lane of a 10x microfluidic chip device. Typically, cells from one organoid were loaded per lane in the microfluidic device, and in some cases organoids from different lines were pooled onto the same lane and demultiplexed based on single-nucleotide polymorphisms. For 1-month organoids, three pooled 409b2 and one H9 organoid were dissociated and cells from the two cell lines were mixed at equal ratios to be loaded on the chip. For a set of 2-month HipSci organoid data, organoids were dissociated for all four HipSci cell lines and pooled at equal ratios to be loaded on one lane of the microfluidic device aiming for 10k cells. Fluidigm C1 data (Supplementary Table 1) were generated as previously described[15] and cells from chimpanzee SandraA 75d organoids were microdissected regions from vibratome slices for which single-cell suspensions were generated as described above. Single cells were then sorted into 96-well plates using a FACS Aria III sorter and further processed using the Smart-seq2 protocol[34] to generate cDNA and the NexteraXT kit (Illumina) to generate sequencing libraries. All libraries (10x and Fluidigm C1/Smart-seq2) were sequenced on Illumina's Hiseq2500 platform in paired-end mode (100 bp Fluidigm C1/Smart-seq2; 26+8 bp, 100 bp 10x).

**Early stages of organoid differentiation (iPSCs to neuroepithelium).** For iPSC/ESC single-cell experiments, cells were detached from cell culture dishes using TrypLExpress (Thermo Fisher) incubation for 5 min, followed by addition of mTeSR1. Cells were centrifuged for 5 min at 200*g* and resuspended in mTeSR1, filtered through a 20-μm strainer and washed with mTeSR1. Cells were then centrifuged again for 5 min at 200*g* and resuspended in mTeSR1, counted, diluted to the same concentration and mixed at equal ratios for the three cell lines to be loaded on the 10x microfluidic chip aiming for 10,000 cells. Thirty embryoid bodies, 7–15 neuroectoderms, and 1–3 neuroepithelia of each cell line were pooled for each dissociation, respectively. Cells were obtained by papain dissociation as described above for organoid dissociation, with slightly shorter incubation times in enzyme mix 1 (approximately 30 min.). For 10x experiments, cells from the three different cell lines were diluted and mixed at equal ratios to be loaded on the microfluidic chip device.

Single-cell experiments were conducted using the 10x Chromium Single Cell 3′ v2 Kit following the manufacturer's instructions. In brief, cells were mixed with reverse transcription mix, gel beads and oil were loaded on the chip device to be coencapsulated into droplets, which underwent first strand cDNA synthesis thereby tagging mRNAs with a UMI and a unique cell barcode. All following steps were conducted in bulk by breaking the droplets and cleaning up and amplifying the cDNA. Single-cell libraries were then constructed by fragmentation, end repair and adaptor ligation and amplification using library-specific index sequences as provided by 10x Genomics. Quantification and quality control of libraries was performed using High Sensitivity DNA assays for Agilent's Bioanalyzer and sequenced on a HiSeq2500 in Rapid or HighOutput sequencing mode. Typically, one 10x library was sequenced on one lane of a sequencing flow cell, with the exception of the HipSci organoids, for which three pooled libraries (each library contained pooled cells from four dissociated HipSci organoids from different cell lines) were sequenced on two lanes of a flow cell. See Supplementary Table 1 for more details.

**Immunohistochemistry.** Organoids were washed in PBS before fixing in 4% PFA for 2–4 h. The excess of fixative was removed with three PBS washes and organoids were then transferred to a 30% sucrose solution for 24–48 h for cryoprotection. Finally, organoids were transferred to plastic cryomolds (Tissue Tek) and embedded in OCT compound 4583 (Tissue Tek) for snap-freezing on dry ice. For immunohistochemical stainings, organoids were sectioned in slices of 20-μm thickness using a Leica CM3050 S cryostat and Microm HM 560 (Thermo Fisher) at −15 to −20 °C. Organoid sections were quickly washed in PBS to remove any residual OCT. Then, sections were incubated in antigen retrieval solution (HistoVT One, Nacalai Tesque) at 70 °C for 20 min. Excess solution was washed away with PBS and the tissue was incubated in blocking-permeabilizing solution (0.3% Triton, 0.2% Tween-20 and 5% normal goat serum in PBS) for 1 h at room temperature. Afterwards, sections were incubated overnight at 4 °C in blocking-permeabilizing solution containing antibodies anti-PAX6 (mouse, 1:1,000, Thermo Fisher, MA1-109; rabbit, 1:300, Covance, PRB-278P) and anti-CTIP2 (rat, 1:1,000, Abcam, AB18465), anti-SATB2 (rabbit, 1:500, Abcam, Ab92446; mouse, 1:500, Abcam, Ab51502), anti-Tbr2 (mouse, 1:500, MPI-CBG Antibody Facility[26]). On the next day, sections were rinsed three times in PBS before incubation for 1 h at room temperature in secondary antibody solution, which contained blocking-permeabilizing solution, DAPI (1:3000), Alexa Fluor 488-conjugated anti-rabbit antibody (goat, 1:1,000, Thermo Fisher, A11008), Alexa Fluor 546-conjugated anti-mouse antibody (goat, 1:500, Thermo Fisher, A-21123), Alexa Fluor 647-conjugated anti-rat antibody (goat, 1:500, Thermo Fisher, A-21247) and Alexa Fluor 488-conjugated anti-mouse (A21202) and anti-rat antibody (A21208), Alexa Fluor 555-conjugated anti-rabbit antibody (A31572), Alexa Fluor 647-conjugated anti-mouse antibody (A31571) (all donkey-derived, 1:500, Molecular Probes). Finally, the remainders of secondary antibody solution were washed off with PBS before covering with ProLong Gold Antifade Mountant medium (Thermo Fisher). Stained organoid cryosections were imaged using a confocal laser scanning Olympus Fluoview FV1200 microscope and Zeiss LSM 880 Airy upright microscope. Whole-section tilescans composed of 3 different $z$-plane images ($z$-step = 5–8 μm) were acquired using a 10× magnification objective, Plan-Apochromat 10×/0.45 M27 and Plan-Apochromat 20×/0.8 M27 objectives. Images were then stitched, stacked and further processed using the Olympus Fluoview 4.2b software and ImageJ (Fiji).

**Single-cell RNA-seq data preprocessing.** We used Cell Ranger, the set of analysis pipelines suggested by 10x Genomics, to demultiplex raw base call files to FASTQ files and align reads to the human genome and transcriptome (hg38, provided by 10x Genomics) with the default alignment parameters. Pooled samples, including samples from different species or human lines, were then demultiplexed using a two-step procedure based on the read mapping results. In the first step, the genome alignment between human (hg38) and chimpanzee (panTro5) was downloaded from UCSC Genome Browser. Sites with diverged bases between human and chimpanzee were obtained based on the genome alignment. Reads covering the species-diverged sites were collected for each reported cell, with the number of bases matching each species counted. Cells with more than 80% reads covering the species-diverged sites matching with one species were assigned as cells from the species. For those samples with human cells from different lines pooled, a second step of demultiplexing was done using demuxlet[35], based on the genotyping information of lines downloaded from HipSci websites (Kucg2, Wibj2, Hoik1, Sojd3) or called using bcftools based on the unpooled scRNA-seq data (H9, 409b2). Cells with the best singlet likelihood no less than 50 higher than the second best singlet likelihood and estimated mixture ratio less than 30% were labelled as their best-matched lines. All cells failing to pass any of the above threshold were classified as doublets and excluded from the following analysis.

Seurat[36] was then applied for further data processing. Cells with more than 6,000 or less than 200 detected genes, as well as those with mitochondrial transcripts proportion higher than 5% were excluded. After the log-normalization, confounding factors including the number of detected genes and proportions of mitochondrial transcripts were also regressed out. Highly variable genes were then obtained as genes with dispersion higher than 0.5 and normalized expression level between 0.0125 and 3, followed by principal component analysis (PCA) based on the $z$-transformed expression levels of the identified highly variable genes (Supplementary Table 2). The top-20 principal components (PCs) were used to do clustering (with a resolution of 0.6) using Seurat. Additional quality controls of the measured cells were based on primary cell-type predictions by using public human fetal brain scRNA-seq data (Nowakowski dataset)[19]. In brief, a Lasso logistic regression model was built, using gene-expression ranks of the Nowakowski dataset as the training set, to predict the primary cell-type identity of each single cell in two-month-old and four-month-old organoids. Cells which were predicted to be of 'glycolysis' identity were excluded, as well as cells in the Seurat clusters in which more than 80% of cells were predicted to be of 'glycolysis' identity. Heterogeneity analysis of human (Extended Data Fig. 2, Supplementary Tables 3, 4) and chimpanzee (Extended Data Fig. 4, Supplementary Tables 5, 6) full lineage data was performed using $t$-SNE based on the top PCs identified (top 20 PCs for human, top 15 PCs for chimpanzee). Cluster identities were assigned based on cluster gene markers (Supplementary Tables 4, 6) as determined by FindAllMarkers function in Seurat (min percentage of cells expressed = 0.25 and log fold change threshold = 0.25) and gene expression of known marker genes. For human data, cells from 409b2 and H9 were integrated using canonical correlation analysis (CCA) as implemented in Seurat (v.3). In brief, data were normalized and the top 2,000 highly variable genes for 409b2 and H9 cells were determined using the vst method. The datasets were integrated based on the top 20 CCs using the Seurat method by identifying anchors and integrating the datasets. The resulting integrated data were scaled and PCA was performed. Clustering was performed based on the top 20 PCs and using a resolution of 0.6. Feature plots show non-integrated expression values. Cluster markers were determined using a Wilcoxon test, considering only genes that show a minimum log fold expression change of 0.25 in at least a fraction of 0.25 of cells in the clusters using the non-integrated expression values.

**RSS and construction of pseudocell transcriptomes.** RSS of one cell to the Human Developing Brain atlas (BrainSpan) was defined as the normalized similarity between gene expression levels of the cell and gene expression levels of each of the 237 fetal samples with RNA-seq data in the BrainSpan database in Allen Brain Atlas. To increase discrimination of different reference samples, only the highly variable genes of the BrainSpan dataset (see Supplementary Table 2), defined based on expression variation-mean comparison of the reference dataset, were used for the RSS calculation. Between each cell and each sample in the BrainSpan dataset, the Pearson correlation coefficient (PCC) was calculated across the BrainSpan highly variable genes. $z$-Transformation was then applied to PCCs between each cell to the 237 fetal BrainSpan samples to get the normalized similarities.

To construct pseudocells, single cells were first grouped based on their sample sources and Seurat clusters. Within each group of cells, that is, those cells from the same sample and in the same Seurat cluster, cells were selected randomly with a selection probability of 20%. The selected cells were called pseudocell seeds or territory capitals. The ten nearest neighbours of each seed, based on Euclidean distances of the top-20 PCs, were then assigned to the seed, forming a pseudocell territory. If one cell was assigned to multiple pseudocell territories, one territory was chosen randomly. The expression level of one gene in each pseudocell was then calculated as the average gene expression level across cells in the pseudocell territory.

**Visualization, lineage identification and pseudotime estimation of pseudocells for reconstructing human cerebral organoid differentiation from pluripotency.** First, PCA was applied to a pseudocell expression matrix using the $z$-transformed expression levels of the highly variable genes as input. Euclidean distance between the top 10 PCs of each pair of pseudocells was calculated and a $k$-nearest neighbour ($k$NN) network ($k = 100$) was then calculated with the constraint to only consider pseudocells from the same or nearby stages when screening for nearest neighbours. The $k$NN network was visualized using SPRING[17]. To construct the pseudotime course of human cerebral organoid differentation from pluripotency, the Walktrap community identification algorithm (implemented in the R package igraph) was applied to the above $k$NN network to identify network communities. The resulting communities were manually aggregated into four groups to minimize branches in each group. A diffusion map algorithm (implemented in the R package destiny[37]) was applied to pseudocells in each of the four groups, with the expression levels of the highly variable genes of pseudocells as the input. The ranks in DC1 were used as the pseudotimes. We used an $F$-test-based ANOVA analysis to identify genes with pseudotime-dependent expression patterns. In brief, we established a natural splined linear regression model (ns function in the R package splines) with six degrees of freedom (df), with expression levels as the response variable and pseudotimes as the independent variable, for each of the highly variable genes. An $F$-test was applied, to compare variation explained by the splined linear model with that of the residuals normalized by degrees of freedom. Bonferroni correction was performed across tested genes, with a corrected $P$ value threshold of 0.01 to identify genes with pseudotime-dependent expression. The analysis was applied to the four groups of pseudocells separately.

**Visualization, lineage identification and pseudotime estimation of cells in human two-month-old cerebral organoids from different individuals.** Pseudocells were constructed for the human two-month-old organoids as above with constraint on samples and based on cells with predicted primary cell types as one of radial glia, intermediate progenitor, excitatory neuron, and inhibitory neuron. RSS to the BrainSpan fetal samples was calculated for each pseudocell, with distance between two pseudocells defined as the correlation distance between RSS of the two pseudocells. The $k$NN network ($k = 20$) was then constructed and SPRING was used to determine coordinates of pseudocells for visualization. To further discriminate pseudocells representing different neuronal lineages, a Walktrap algorithm for network community identification was applied to the RSS-based $k$NN ($k = 100$). Communities that were significantly connected and showing concordant marker expression or similarity spectrum were aggregated, which resulted in three progenitor-to-neuron trajectories. Based on gene expression level ranks of cells in the three defined trajectories, two Lasso logistic regression models were trained, one for classification of cortical and ventral lineage and the other for classification of all the three trajectories. The first model was applied to pseudocells in the community C6, which was significantly connected to cortical and ventral trajectories, while the second model was applied to pseudocells in the community C4, which was significantly connected with both the non-telencephalon pseudocells and community C4. With a unique lineage label defined for each pseudocell, a $1 \times 500$ self-organizing-map (SOM) model was trained for each of the three trajectories, using RSS of pseudocells within the lineage as the training data. The index of neuron to which one pseudocell was assigned was used as its pseudotime. Diffusion map analysis was also applied to pseudocells at each trajectory, with highly variable gene expression as the input, with ranks of DC1 defined as alternative pseudotime of pseudocells. Pseudotimes obtained by the two methods are highly correlated (Spearman correlation is 0.91 and 0.92 for the dorsal and ventral telencephalon trajectories, respectively).

To project the single-cell data to the cell embedding space that was defined for pseudocells, two support vector regression (SVR) models (implemented in the R package e1071), each of which was for one dimension of the embedding, were trained using RSS of pseudocells as the training set. The trained models were applied to RSS of single cells for their predicted coordinates. Such coordinates were further refined by pushing each cell to its nearest pseudocell with smallest correlation distance of RSS to be 70% closer. Similarly, a support vector machine (SVM) model was trained (implemented in the R package e1071) using RSS of pseudocells for the three trajectories, and applied to RSS of single cells for their trajectory identity. After that, the corresponding SOM model for pseudotime estimation was applied to RSS of each single cell for its estimated pseudotime.

**Dynamic time warping-based alignment of pseudotime courses.** We used a dynamic time warping (DTW) algorithm to align different pseudotime courses. In brief, each pseudotime course was evenly broken into 50 blocks. Average gene expression levels of pseudocells or cells within each block was calculated. Pairwise distances between blocks from the two courses were calculated as the Pearson correlation distance, that is, $1 - \mathrm{PCC}$, across the highly variable genes in cells of both pseudotime courses. Suppose $d_{i,j}$ represents the distance between the $i$-th block in the reference pseudotime course and the $j$-th block in the query pseudotime course. We defined $\mathbf{D}$ as the alignment distance matrix, where

$$\mathbf{D}_{i,j} = \min\left(\mathbf{D}_{i-1,j}, \mathbf{D}_{i,j-1}, \mathbf{D}_{i-1,j-1}\right) + d_{i,j}$$

A trace-back procedure was then performed to get the alignment. Three modes of alignment were implemented. In the first mode, the fixed-end alignment, the initialization of $\mathbf{D}$ was done as:

$$\mathbf{D}_{i,j} = \begin{cases} d_{i,1} & \text{if} \quad j = 1 \\ \mathbf{D}_{1,j-1} + d_{1,j} & \text{if} \quad i = 1 \end{cases}$$

In the other two modes, the fixed-start and end-to-end alignments, **D** was initialized as:

$$\mathbf{D}_{i,j}=\begin{cases} d_{1,1} & \text{if} \quad i=1 \quad \text{and} \quad j=1 \\ \mathbf{D}_{1,j-1}+d_{1,j} & \text{if} \quad i=1 \\ \mathbf{D}_{i-1,1}+d_{i,1} & \text{if} \quad j=1 \end{cases}$$

In the trace-back step, a fixed-end and end-to-end alignment was started from $\mathbf{D}_{M,N}$, where $M$ and $N$ are the numbers of blocks at the reference and query pseudotime courses, respectively. On the other hand, the trace-back step was started from $\mathbf{D}_{m,N}$, where $m = \text{argmin}_i(\mathbf{D}_{i,N})$. In our study, the fixed-end alignment was used to align the cortical and ventral lineage pseudotime course of human organoid cells; the fixed-start alignment was used to align pseudotime courses of human and chimpanzee cortical pseudocells; the end-to-end was used in the truncated alignment of pseudotime courses of different species.

**Reconstruction of chimpanzee cerebral organoid differentiation from pluripotency.** We applied a similar procedure to that mentioned above describing the reconstruction of human cerebral organoid differentiation from pluripotency to reconstruct the organoid differentiation trajectory from chimpanzee single-cell RNA-seq data. In brief, the scRNA-seq reads were mapped to the human-chimpanzee-macaque consensus genome and counted using Cell Ranger. Seurat was used for further preprocessing including gene expression normalization, confounding factor regression, PCA and clustering. Cells from organoid samples with predicted primary cell-type identity of 'glycolysis', as well as cells within clusters with more than 80% cells having 'glycolysis' identity, were excluded. Pseudocells were then constructed with a seed selection probability of 20% and constraints on samples and Seurat clusters. PCA was applied to expression levels of highly variable genes across pseudocells, and pairwise distances of pseudocells were calculated as the Euclidean distances between the top-10 PCs. The $k$NN network ($k = 100$) of pseudocells was constructed, linking every pseudocell with its 100-nearest pseudocells representing the same or nearby stages. Three-month-old and four-month-old organoids were seen as the same stage. The Walktrap network community identification algorithm was applied and the resulted community labels (walktrap communities) of pseudocells were compared with the predicted community labels (projected communities) based on a Lasso logistic regression model trained by ranks of gene expression levels of the human pseudocells representing the human organoid differentiation from pluripotency as described above. Any walktrap community with <1,000 $k$NN connections with other communities was discarded. One of the four labels: early, cortical, ventral, non-telencephalon was assigned to one walktrap community if more than 95% of pseudocells within the community were from the same group according to their projected communities. For one community with more than 10% of pseudocells with projected communities belonging to both ventral and midbrain-hindbrain groups, the non-telencephalon label was only assigned to pseudocells with projected communities in the non-telencephalon group. The diffusion map algorithm was applied to each of the four pseudocell groups, using the expression levels of highly variable genes as input, to estimate their pseudotimes. For the cortical, ventral and midbrain-hindbrain groups, the ranks of DC1 were used as the pseudotimes. For the early group, a principal curve (implemented in the R package princurve) was fitted in the DC1–DC2 space. The order of pseudocells projecting to the resulted principal curve was used as the pseudotime.

**Human-chimpanzee-macaque consensus genome.** The construction of the consensus genome was performed as described[3,38]. In brief, the chained and netted pairwise genome alignments of the human (hg38) and chimpanzee (panTro5) genomes, and the human and macaque (rheMac8) genomes, were downloaded from UCSC Genome Browser. Based on the downloaded pairwise genome alignments, a multiple genome alignment of human-chimpanzee-macaque was constructed using multiz. On the basis of the human-chimpanzee-macaque genome alignment, we constructed the consensus genome by masking all discordant sites including mismatches, insertions/deletions (indels), as well as 6-bp flanking regions of indels on the human genome. The obtained consensus genome was indexed with gene annotation in GENCODE v.27 for read mapping to the consensus genome with Cell Ranger.

**Pseudotime estimation of cerebral organoid cells in different species.** Single-cell RNA-seq data of organoids with ages from two months to four months in human, chimpanzee and macaque were mapped to the human-chimpanzee-macaque consensus genome and counted using Cell Ranger. Further preprocessing using Seurat was applied separately for data from the three species. Only cells with predicted primary cell-type identities as radial glia, intermediate progenitors, excitatory neurons or inhibitory neurons were included in the later analysis. Pseudocells were constructed for humans and chimpanzees, both with a coarse grain ratio of 20% and constraints on samples and Seurat clusters. The RSS to the BrainSpan dataset was calculated for each pseudocell, and the SVM model for lineage estimation was applied to estimate the lineage identity of each pseudocell. Focusing on the cortical lineage, a diffusion map analysis was applied to cortical pseudocells of the three species, respectively. The ranks of DC1 were used as the pseudotimes of the pseudocells. In macaque, a similar procedure was applied directly to single cells without pseudocell construction.

**Truncated DTW-based alignment of pseudotime courses representing neural progenitors and deeper-layer neurons in different species.** We used the first DC discriminating *BCL11B*$^+$ and *SATB2*$^+$ cortical neurons (DC3 in chimpanzee, DC4 in macaque) to identify upper-layer (UL) neurons, as the pseudocells in the branch with highest expression level of *SATB2*. To identify potential UL neurons in human, we first retrieved markers of upper and deeper-layer (DL) excitatory neurons[19]. The sum expression levels of UL and DL markers was then calculated for each pseudocell in human and chimp, with the UL-specificity score ($s_{\mathrm{UL}}$) being defined as the expression ratio of UL/DL markers. The distribution of $s_{\mathrm{UL}}$ in UL neurons in chimpanzee was used to determine the threshold to discriminate UL neurons from other cell types ($s_{\mathrm{UL}} > 0$). All UL neurons in the three species were excluded from the following analysis.

To correct for the DL neuron maturation timing differences between human and the other two species, a two-step pseudotime course alignment strategy was used. The first step, namely the trimming step, aims to determine the pseudotime points in chimpanzee and macaque which correspond to the latest pseudotime point in human. In brief, an SVR model with Gaussian-kernel was first constructed, with chimpanzee or macaque pseudotimes as the response variables and the RSS as the dependent variables. Two models were trained with the chimpanzee pseudocells and macaque cells, respectively, and applied to the human pseudocells to predict their corresponding chimpanzee and macaque pseudotime points. Two constrained B-splines regression models ($F_{\mathrm{HC}}$, $F_{\mathrm{HM}}$) were then fitted (implemented in the R package cobs): human pseudotimes of human pseudocells ($t_{\mathrm{h}}$) versus their predicted chimpanzee $(\widehat{t_{\mathrm{c}}})$ or macaque $(\widehat{t_{\mathrm{m}}})$ pseudotimes, with constraints of $F_{\mathrm{HC}}(t_{\mathrm{h}} = 0) = F_{\mathrm{HM}}(t_{\mathrm{h}} = 0) = 0$. $F_{\mathrm{HC}}(t_{\mathrm{h}} = 1)$ and $F_{\mathrm{HM}}(t_{\mathrm{h}} = 1)$ were used as the pseudotime thresholds to select chimpanzee pseudocells and macaque cells. Chimpanzee DL neurons with pseudotime $t_{\mathrm{C}} > F_{\mathrm{HC}}(t_{\mathrm{h}} = 1)$, as well as macaque DL neurons with pseudotime $t_{\mathrm{M}} > F_{\mathrm{HM}}(t_{\mathrm{h}} = 1)$, were excluded in the following analysis. The second step, namely the alignment step, was then applied to the remaining pseudotime courses of the three species. An end-to-end DTW-based alignment, as described above, was used to align the human pseudotimes with pseudotimes of each of the other two species using the human pseudotime course as the template.

**Identification, clustering and species specificity of differentially expressed genes between humans and chimpanzees.** Human genes resulting from duplication or rearrangement that do not exist in other apes were collected via previous studies[26,39–41]. In total, 41 genes were obtained, 24 of which were included in the human gene annotation in Cell Ranger. 23 of them were detected in at least one human cell, with *FCGR1B* being the exception. Expression patterns of those genes were examined and summarized manually (Supplementary Table 7).

To compare transcriptome changes of the developmental trajectory from cortical neural progenitors to DL neurons between human and chimpanzee, an $F$-test-based comparison was applied to the expression profile along pseudotimes of the two species. In brief, for each gene, a natural spline linear regression model (df = 6) was constructed for human and chimpanzee pseudocells along the aligned pseudotime course, without discriminating human and chimpanzee samples, and used as the null model ($m_0$). The alternative natural spline linear regression model was also constructed, with each species having its own slopes and intercept ($m_1$). The residuals of the variation, which cannot be explained by each model, were compared by an $F$-test. Non-ribosomal genes with BH-corrected $P < 0.01$ were identified as differentially expressed (DE) genes between human and chimpanzee along the developmental trajectory from cortical neural progenitors to DL neurons (Supplementary Table 8).

To estimate the robustness of the identified differential expression to the number of used lines, as well as the pseudocell distribution along the pseudotime course, we used a series of replaceable pseudocell sampling procedures with constraints. In brief, in each round of replaceable pseudocell sampling, the candidate pseudocells to be selected are restricted to be those from a certain number of human cell lines. In addition, the subsampling in human pseudocells is performed to recapitulate the pseudocell distribution along the aligned pseudotime of chimpanzee pseudocells; that is, each of the ten pseudotime bins contains the same number of human and chimpanzee pseudocells. This sampling procedure was performed 100 times for each possible number of human lines, ranging from one to seven. Differential expression analysis, as described above, was applied to compare gene expressions of human pseudocells in each sampling with the chimpanzee pseudocells. Robust DE genes were determined as DE genes which can be detected in at least 80% of tests performed with replaceable pseudocell samplings with any number of used human cell lines.

A similar strategy was also used to estimate the false-positive human–chimpanzee DE genes due to differences between cell lines. In each sampling, two lines were randomly selected as group one, and a certain number of lines, ranging from one to five, were selected from the remaining lines as group two. For each group, pseudocells were randomly sampled from the selected lines to recapitulate the pseudocell distribution along the aligned pseudotime of chimpanzee pseudocells. Such sampling was performed 100 times for each possible number of lines used in group two. The transcriptome trajectory from cortical neural progenitors to DL neurons in macaque organoids was used as the evolutionary outgroup to determine species specificity of the identified human–chimpanzee DE genes. First, the cumulative expression divergences of each gene between human and macaque ($d_{HM}$), and between chimpanzee and macaque ($d_{CM}$), were calculated. The cumulative expression divergence was calculated by summing up absolute values of average expression differences between species at the 50 pseudotime bins of equal sizes along the aligned pseudotimes. The human–chimpanzee DE of one gene is seen as human-specific if $d_{HM} - d_{CM} > \max(d_{HM}, d_{CM})/2$. Genes with chimpanzee-specific differential expression were identified in the same way. Genes with human-specific differential expression were then clustered based on their human–chimpanzee DE along pseudotimes. Average expression differences between human and chimpanzee at each of the 50 pseudotime bins along the pseudotimes was calculated for each gene with human-specific differential expression (denoted as $d_t$ at pseudotime bin $t$), and then normalized as $\widehat{d_t} = d_t / \left(\max_t d_t - \min_t d_t\right)$. Hierarchical clustering (Ward algorithm) was then used to cluster those genes into nine clusters, with distances between genes calculated as the Euclidean distances between their normalized differential expression spectrums. Clusters with fewer than five genes were discarded. We annotated genes with human-specific expression patterns using the *Homo sapiens* Gene Ontology Annotation file (validation date: 21 April 2017) provided by the Gene Ontology Consortium.

**Processing of the Fluidigm C1-based scRNA-seq data of cerebral organoids.** In addition to the newly generated Fludigm C1 (Smart-seq2)-based scRNA-seq data, we further retrieved published sequencing data of 786 and 344 single cells from human and chimpanzee cerebral organoids[11,15], in the format of FASTQ files from GEO accession numbers GSE75140 and GSE86207 (CMK dataset). All the reads were mapped to the human-chimpanzee-macaque consensus genome using STAR (v.2.6.1d) with '–quantMode' parameter set to TranscriptomeSAM and GENCODE v.27 annotation provided. Gene expression levels in each cell were quantified as TPM by RSEM (v.1.3.1). Additionally, we retrieved the recently published gene-expression matrix representing 3,211 cells from human and chimpanzee cerebral organoids (excluding redundant cells from GSE75140 and GSE86207) and 4,854 cells from human and macaque fetal brains[16].

Based on the resulting gene-expression profile, RSS to the fetal Brainspan dataset was calculated as described above for each cell, with 248 genes with significant differential expression between cortical neurons measured by Smart-seq and Smart-seq2 excluded from the references. Distances between organoid cells were calculated as the Pearson's coefficient distances between RSS of cells. Distances between cells from fetal brains were calculated in the same way. The resulting distance matrices of all organoid cells and fetal brain cells were used as the input to generate *t*-SNE embeddings. A *k*NN network ($k = 50$) was generated for organoid cells and fetal brain cells separated based on the RSS-based distances, and a Walktrap algorithm for network community identification was applied to identify cell clusters, which were further annotated based on their marker genes. Based on the cell-type annotation, the diffusion map analysis, with the RSS profiles as input, was applied to the dorsal forebrain NPCs and neurons in organoids and fetal brains, respectively. The ranks of DC1 were used as the pseudotimes.

To validate the human–chimpanzee differential expression identified in our droplet-based scRNA-seq data using the C1-based cerebral organoid data, the organoid dorsal telencephalon pseudotemporal trajectory was first split into ten intervals. In each pseudotemporal interval, the human–chimpanzee DE was calculated as the $\log_2$-transformed fold change ($\log_2$FC) between the average expression of human and chimpanzee cells in the interval. Here, the CMK dataset and other datasets that used a distinct quantification method were processed separately. A similar strategy was also applied to the aligned droplet-based human and chimpanzee pseudotemporal trajectories. Generalized log2-transformed fold change (g$\log_2$FC), defined as the average $\log_2$FC across the pseudotemporal intervals, as well as the maximum $\log_2$FC across the intervals (m$\log_2$FC), was further calculated for each human–chimpanzee robust DE genes in organoids. A DE gene is seen as being consistent in the two datasets if both g$\log_2$FC and m$\log_2$FC of the C1-based and droplet-based human–chimpanzee comparisons are of the same signs (referred to as consistent DE genes). The pseudotemporal intervals with the maximum fold change in the droplet-based and C1-based trajectories were also obtained and compared for the consistent DE genes. This procedure was also applied to compare human–macaque differential expression of the human-specific DE genes along the droplet-based pseudotemporal trajectory and the C1-based fetal brain pseudotemporal trajectory.

**Single-cell and bulk ATAC-seq data generation.** Two-month-old and four-month-old organoids were washed twice with PBS in a Tissue-Tek Cryomold (Sakura), then embedded in 4% low-melting agarose (Sigma) and sliced into 150-μm sections using a vibrating microtome (Ci 7000 smz, Camden Instruments). Slices were placed on microscope slides containing differentiation medium with vitamin A (Diff+VA) and inspected under a stereomicroscope to dissect cortical regions. Selected regions were washed twice in 500 μL PBS and incubated at 37 °C in 500 μL Accutase (Sigma) plus 0.5 μL DNase I (New England Biolabs) for ~45 min. Trituration was performed for additional mechanical dissociation. Cells were passed through a 30-μm pre-separation filter (Miltenyi Biotec), washed with Diff+VA medium, and spun down at 300*g* (Heraeus Megafuge 40R, Thermo Fisher) for 5 min. The cell pellet was resuspended in 200 μl of Diff+VA medium. Cells were viewed under a microscope to ensure a single cell suspension was obtained, and then counted using a Countess Automated Cell Counter (Invitrogen). Single cell suspensions for the early stages of organoid differentiation (iPS cells to neuroepithelium) were obtained as described above.

From the cell suspension, 50,000 cells were used as input for bulk ATAC-seq as described[42]. The remaining cells were diluted to a final concentration of 300 cells/μl and used for microfluidics-based single-cell ATAC-seq as described[43]. In brief, cells were mixed with Suspension Reagent (Fludigm) at a 3:2 ratio and loaded onto a primed medium (10–17 μm) integrated microfluidic circuit (Fludigm) for capturing. Cell-capture sites were examined under a microscope and noted for containing 0, 1, or multiple cells. Lysis, transposition, and amplification were performed on the Fluidigm C1 platform. DNA from each cell was transferred to an individual well of a 96-well plate and barcoded with unique combinations of 24 adaptor-index i7 and 16 adaptor-index i5 primers[43]. Quantification and library size distribution was assessed on an Agilent 2100 Bioanalyzer using High Sensitivity DNA chips. Excessive primer contamination was removed using SPRIselect (Beckman Coulter Life Sciences) size selection. Up to 192 cells were pooled and sequenced in paired-end, dual-index mode for 50 + 8 + 50 + 8 cycles on one lane of an Illumina HiSeq 2500. A summary of all single-cell experiments can be found in Supplementary Table 1.

**Single-cell and bulk ATAC-seq data processing.** Base calling was performed using Bustard (Illumina), adaptor trimming with leeHom[44] and demultiplexing with deML[45]. Reads were aligned to hg19 for human, panTro4 for chimp and rheMac8 for macaque using bowtie2 with a maximum fragment length of 2,000. PCR duplicates were marked and removed using Picard tools (http://broadinstitute.github.io/picard). Samtools[46] was used to retain properly paired reads with mapping quality greater than 30, while reads mapping to the mitochondrial genome, Y chromosome and blacklisted genomic regions that show excessively high read mapping, several of which correspond to nuclear mitochondrial DNA segments (identified in ref. [43] and the ENCODE Project[47]) were removed. For scATAC-seq, single cell BAM files were merged, excluding data from any capture site with 0 or more than 1 cell, to create an aggregated BAM file. Peaks, which represent regions enriched in mapped pair-end sequences, were called using MACS2[48] with options nomodel, nolambda, keep-dup all and call-summits. Peak summits were extended by ±250 bp. In the event of overlapping peaks, the peak with the lowest *P* value was kept. A single-cell ATAC-seq consensus peak set was obtained by requiring a peak to be accessible in a minimum of 5% of cells. Data visualization was carried out using the Integrative Genomics Viewer (IGV)[49].

**Enrichment for validated human VISTA enhancers.** We overlapped scATAC-seq peaks detected in human cerebral organoids with positive human VISTA enhancers using bedtools intersect. For each tissue annotated in the VISTA Enhancer Browser, we counted the number of enhancers that did or did not overlap a peak. We compared these values to the number of all other tissue elements that did or did not overlap a peak. Fisher's exact tests were performed to determine which tissues' enhancers had a higher likelihood of being represented. The significance values were corrected for multiple testing using the qvalue package in R.

**Cell-state identification using single-cell ATAC-seq on cerebral organoids and pseudotime estimation.** The accessibility at each site in the consensus peak set for every single cell was used to create a count matrix. Cells with fewer than 5,000 read pairs and less than 5% of reads in peaks (fraction of reads in peaks, FRiP) were filtered out from further analyses. chromVAR[29] was used to scan the peaks for transcription factor binding motif occurrences, using a curated collection of 1,765 human motifs from the cisBP database, and to identify significantly variable motifs among cells. In addition to transcription factor (TF) binding motifs, peaks were scanned for 7-mers. Cell similarity was visualized in a two-dimensional *t*-SNE plot using the bias-corrected deviations in accessibility for 7-mers.

Each cell's *t*-SNE coordinates and the consensus peaks were passed to Cicero[50] and the densityPeak algorithm was used to identify two clusters of cells. Statistically significant differences in TF motif accessibility between the two clusters was calculated using chromVAR, and those motifs corresponding to marker transcription factors known to distinguish neural progenitors and neurons were used as the basis for cell-state identification. Statistically significant differences in accessibility of additional annotations between the two clusters were used to support cell-state identities. These annotations included differentially accessible chromatin peaks identified as being enriched in developing mouse brain radial glial cells or excitatory neurons[51], as well as accessibility in peaks near to genes showing pseudotime-dependent expression in cortical neural progenitors or cortical neurons identified as part of this study.

We identified differentially accessible (DA) peaks between the two clusters using the command differentialGeneTest in Cicero. A count matrix was generated with featureCounts[52] using the top 250 DA peaks in each cluster. This count matrix was used as input for a diffusion map in order to obtain a pseudotemporal ordering of the cells[53]. Projecting transcription factor binding motif deviation $Z$-scores on the cells revealed a gradient of known neural progenitor to neuronal markers along the first diffusion map component and we took a cell's rank along this component as its pseudotime value.

DA peaks identified between the two clusters were used as input test regions for GREAT (v.3.0.0)[54] with all accessible organoid peaks serving as background regions. We used the default basal plus extension genomic association rule with its default values. All GO Biological Process terms and their associated hypergeometric $P$ values were exported. For each term, we plotted the $P$ value obtained using cluster 1 DA peaks and the $P$ value obtained using cluster 2 DA peaks. Terms with $P < 0.05$ were considered enriched. Informative enriched terms were highlighted based on their significance value in one cell state relative to the other, and for small differences between the cell states when highlighting terms enriched in both.

**Single-cell ATAC-seq pseudotime estimation for cells in early states of differentiation and cerebral organoids.** Similar to the analysis of the cerebral organoids, we used chromVAR to calculate bias-corrected deviations in accessibility for TF motifs and 7-mers for each cell across the differentiation trajectory. Here, we included the scATAC-seq consensus peak sets called in the iPSC, embryoid body, neuroectoderm and neuroepithelial time points, in addition to the scATAC-seq consensus peak set from the cerebral organoid time point. In the event of overlapping peaks, the peak with strongest signal was retained. Cells with fewer than 5,000 read pairs and less than 5% of reads in peaks were removed from further analyses (Supplementary Table 9). Cell similarity was visualized in a two-dimensional $t$-SNE plot using the bias-corrected deviations in accessibility for 7-mers.

As the cerebral organoid cells' pseudotimes were previously resolved, we focused on ordering the earlier stages. For this we used Cicero's differentialGeneTest to identify DA peaks among the iPSC, embryoid body, neuroectoderm and neuroepithelial time points. A count matrix was generated using the top 250 DA peaks in each time point and used as input for a diffusion map. Projecting TF motif deviation $Z$-scores of the cells revealed a gradient of pluripotent to more-differentiated marker TFs along the first three diffusion map components. We fitted a principle curve through the map, and used the pluripotent cells as a starting point to guide the curve. The rank of a cell along this curve was used as its pseudotime. We then added the cerebral organoid cells' pseudotime ranks to this earlier stage resolved pseudotime. We used the pheatmap R package to visualize the dynamics of significantly variable TF motifs across pseudotime.

**Annotation of accessible chromatin peaks.** Peaks were linked to an expressed protein-coding gene using the nearest (maximum distance 1 Mb) transcription start site (TSS) of the canonical transcript as defined by GENCODE (comprehensive gene annotation, release 19). Promoter regions were defined as 1,000 bp upstream of a TSS, and distal regions refer to non-promoter regions. Exon and intron annotations were also obtained from GENCODE (comprehensive gene annotation, release 19). BEDtools[55] was used to annotate peaks for several evolutionary signatures, including: human accelerated regions[56–58]; selective sweeps compared to great apes[59] and archaic humans[60]; SNCs in modern humans that happened since the split with great apes and before or after the split with the ancestor of Neandertals and Denisovans, first identified in ref. [61] and updated for this analysis using the most current 1,000 Genomes Phase 3 allele frequencies, with a global allele frequency $\geq$ 99.5% defined as fixed in all modern humans; small indels (up to 5 nucleotides) fixed in modern humans that happened since the split with great apes and before or after the split with the ancestor of Neanderthals and Denisovans[62]; and human deletions that are highly conserved in mammals[63] (hCONDELs, Supplementary Table 11).

**Identification of genomic regions with differential accessibility between human and chimpanzee organoid neural progenitors and neurons.** To compare the chromatin accessibility of NPCs and neurons in cerebral organoids between human and chimpanzee and identify putative regulatory regions that may contribute to transcriptome divergence between human and chimpanzee, we applied a likelihood ratio test based on a generalized linear model with binomial error distribution to each regulatory region identified in human and chimpanzee organoids. More specifically, we identified open chromatin regions in human and chimpanzee organoids separately as described above. To compare an equal number of human and chimpanzee regions, we took the top 77,611 peaks (corresponding to the number of human consensus peaks) in each species and performed reciprocal liftOver, requiring a 50% minimum ratio of bases that must remap, in order to identify their orthologous counterparts in the other species. Peaks that successfully lifted over (>99%) were merged using bedtools and renamed (that is, mergePeak#). Count matrices were generated at these merged peaks in the species' own genome, and the matrices were then joined on the common peak name. Considering the higher read coverage in human cells, we subsampled reads in human cells to equalize the medians of total number of reads mapped to the regions of interest in human and chimpanzee. This procedure was applied separately to NPCs and neurons. The resulting count matrices were binarized. We then fitted a generalized linear model for each region across all human and chimpanzee cells, with the accessibility as the response variable and species as the independent variable. Another model with the species variable replaced by a scaling coefficient was also fitted as the null model. The scaling coefficient is fixed to one for human cells and $p_c/p_h$ for chimpanzee cells, where $p_c$ and $p_h$ are the average accessibility across all regions and all cells in chimpanzee and human, respectively. We compared the two models and got the $P$ values by using the likelihood ratio test. Regions with BH-corrected $P < 0.01$ were defined as DA regions (Supplementary Table 10). This procedure was applied to NPCs and neurons separately to obtain DA regions in the two cell states.

**Functional and evolutionary characterization of genomic regions with differential accessibility.** We performed permutations to determine if DA peaks were significantly more likely to overlap a given annotation compared to non-differentially accessible (non-DA) peaks. In more detail, we first resized all peaks to an equal length of 500 bp and calculated the average accessibility of human and chimp cells in the resized DA and non-DA peaks. Peaks were then placed into average accessibility bins of 5% intervals. Given the number of DA peaks in each accessibility bin, the same number of non-DA peaks was chosen at random from the corresponding accessibility bin. The random set of non-DA peaks was then overlapped with the given annotation using bedtools intersect. The random sampling of non-DA peaks and annotation overlap was repeated 2,000 times. For each annotation, we counted the number of times a non-DA peak permutation resulted in a higher overlap than what was observed for DA peaks. This number was divided by the number of permutations to determine significance ($P < 0.05$).

We used fixed SNCs, organoid-specific peaks and linked DE genes as annotations. When overlapping peaks with fixed SNCs, we restricted the analysis to include only regions that passed a stringent genome alignability filter ('map35_100%')[61], in which SNCs could be called. Organoid-specific peaks were defined as peaks detected in 2-month and 4-month-old cerebral organoid stages, but not detected in earlier stages of differentiation (pluripotency to neuroepithelial stages). Cell state-specific peaks were those identified as differentially accessible between NPCs and neurons in either human or chimp.

To study putative effects of fixed SNCs on TF binding in the accessible genomic regions, we used funseq2[64] to scan and statistically evaluate all possible TF binding motifs created by fixed SNCs in DA peaks. To generate a list of TF motifs lost on the human lineage, we used the human allele as the reference allele and the ancestral allele[61] as the alternative allele. To generate a list of TF motifs gained on the human lineage, we flipped the state of the reference and alternative allele. This allowed us to directly compare the sequence scores of TF motifs gained or lost in humans. We subtracted the sequence score with the alternative allele from the sequence score with reference allele and performed min-max normalization. Human TF motif gains were plotted as positive values, while human TF motif losses were plotted as negative values. The genomic location of SNCs predicted to alter TF motif binding are provided in Supplementary Table 11. The alteration rate for TF motifs gained in humans was calculated by dividing the number of gains in DA peaks by the number of occurrences of that motif when scanning all organoid accessible peaks using chromVAR and the human genome sequence. The alteration rate for TF motifs lost in humans was calculated by dividing the number of losses in DA peaks by the number of occurrences of that motif when scanning all organoid accessible peaks using chromVAR and the chimpanzee genome sequence. The alteration rates of human TF gains and losses were also calculated per TF family, using TF motif family assignments obtained from ref. [65].

We used the macaque cerebral organoid scATAC-seq data to determine species specificity of the peaks identified as differentially accessible between human and chimpanzee (Supplementary Table 10). In brief, we counted read coverage of each accessible region that we compared between human and chimp that can lift over to the macaque genome in each macaque cell. Regions that failed during liftover were seen as inaccessible in all macaque cells. A random sampling of reads in human and chimpanzee cells was applied to equalize median read coverage in the three species. This procedure was applied 100 times and to the two cell states separately. Accessible probability was then calculated for the two cell states in the three species. In human and chimpanzee, averages across the 100 read-subsampling-based estimation were used. The difference of accessible probability between human and macaque (H − M), and that between chimpanzee and macaque (C − M), was then calculated for each human–chimpanzee DA peak in each cell state. The identified

DA was considered as human-specific if its H – M difference is at least four times larger than the C – M difference, while its H – M difference is no less than 2%. Similar criteria were also applied to define chimp-specific DA.

To investigate potential biological processes that may be influenced by DA peaks, we used human–chimp DA peaks for each cell state (NPC or neuron) as input test regions for GREAT (v.3.0.0)[54] with all accessible organoids peaks serving as background regions. This analysis was then carried out in the same way as explained above.

**Single-nucleus and bulk RNA-seq data generation.** Tissue cubes were dissected from frozen post-mortem prefrontal cortex tissue from human, chimpanzee, bonobo and macaque individuals. In total, three healthy adult human (*H. sapiens*), two healthy adult chimpanzee (*Pan troglodytes*), one healthy adult bonobo (*Pan paniscus*) and three healthy adult rhesus macaque (*Macaca mulatta*) brains were used. The human samples were obtained from the Chinese Brain Bank Center (CBBC) in Wuhan, China. For each of these individuals, written informed consent to use human tissues for research was obtained either from the donors themselves or from their next of kin. All subjects were classified as normal by forensic pathologists at the brain bank, and suffered sudden deaths with no prolonged agonal state. One chimpanzee sample was obtained from Biomedical Primate Research Centre, Netherlands; the other chimpanzee sample was obtained from Yerkes National Primate Research Center, USA. The bonobo sample was obtained from Lola ya Bonobo Sacturary, Congo. Tissue was shipped to the Max Planck Institute for Evolutionary Anthropology under the institutional permit for the transport of biological material derived from endangered species (DE216-08, http://cites.org/common/reg/si/e-si-beg.shtml). Rhesus macaque samples were all collected at the Primate Research Center in Goettingen, Germany. All the chimpanzee and bonobo individuals suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used. All the macaque individuals were euthanized. The Biomedical Research Ethics Committee of Shanghai Institutes for Biological Sciences reviewed the use and care of the animals in the research project (approval ID: ER-SIBS-260802P).

Dissection was performed on dry ice aiming for cubes with minimal curvature to obtain reproducible slicing results. In brief, the thickness of grey matter at all facets of the cube was measured to obtain a mean grey matter thickness. The mean thickness was divided by 10 to obtain the thickness for each of the segments, whereby each of the segments consisted of several slices at 50-μm thickness. Sectioning was performed in a cryostat (Microm, Thermo Fisher), with slices being alternately immersed in Trizol (Invitrogen) for bulk RNA isolation or transferred to a dry tube (low binding) for single nucleus isolation on dry ice. Segments 11 and 12 were collected as well but were considered as being derived from white matter of the cortex. Samples were then stored at −80 °C until further use.

For nuclei isolation from frozen tissue, all following steps were performed on ice with precooled buffers and centrifugation steps were performed at 4 °C. In brief, tissue was spun down, thawed on ice and 1 ml PBSE (PBS (Gibco), 2 mM EDTA (Life Technologies)) was added to the tissue. The tissue slices were incubated at 4 °C on a shaker at 1,500 r.p.m. for a total of 45–60 min with trituration steps in between using 1,000p and 200p to homogenize the tissue. Generally, segments 1–10 were used for single-nucleus experiments. Two segments were pooled to obtain sufficient material for single nucleus isolation, resulting in 5 segments per individual. To reduce batch effects and increase the number of nuclei per experiment, material from three different individuals (originating from human, chimp/bonobo and macaque, respectively) was pooled for each segment. After homogenization, solutions were combined in a 5-ml tube and spun down at 900*g* for 5 min. The pellet was resuspended in 1.5 ml PBSE + 1% NP-40 (BioVision), triturated 20 times using 1,000p and incubated for 7 min incubation on ice. Samples were then spun down at 900*g* for 5 min and resuspended in 1.5 ml PBSE + 1% BSA (Serva) twice. Samples were then spun down again at 900*g* for 5 min and resuspended in PBS + 1% BSA. Before sorting, samples were filtered through a 30-μm cell filter (Miltenyi Biotec) and stained using DAPI (1:1,000, BD Pharmingen). Nuclei were sorted in yield sort mode (BD FACS AriaIII and BD FACS Fusion) based on a defined nuclei population by excluding debris using FSC and SSC and by sorting DAPI positive events. Nuclei were sorted in bulk into 96-well plates and spun down for 5 min at 600*g* to enrich for nuclei in the pellet.

For each of the pooled samples, two lanes on a 10x Chromium microfluidic chip were loaded, aiming for the maximum possible number of nuclei to be targeted obtained from the sorting. Single-nucleus experiments were performed using the 10x Genomics Single Cell 3′ kit v.2 to encapsulate nuclei along with barcode tagged beads, generate and amplify cDNA and to generate sequencing libraries. Each pooled library was barcoded using i7 barcodes provided by 10x Genomics. cDNA and sequencing library quality and quantity were determined using Agilent's High Sensitivity DNA Assay. Libraries were pooled and sequenced in 150-bp paired-end mode on Illumina's NovaSeq platform as provided in Supplementary Table 1.

RNA isolation for bulk RNA-seq was performed using the Direct-zol 96 RNA kit (Zymo Research) and was quantified using Agilent's Bioanalyzer RNA 6000 Nano and Pico kit. Libraries were prepared using the NEBNext Ultra Low RNA Library Prep Kit (New England Biolabs). Library quantification was performed using Agilent's Bioanalyzer DNA 1000 chip kit. All bulk RNA-seq libraries were pooled at equal ratios and sequenced on one lane of an Illumina NovaSeq platform in 150-bp paired-end mode.

**Processing of single-nucleus and bulk RNA-seq data from human, chimpanzee and macaque adult brains.** Single-nucleus libraries were demultiplexed based on their i7 index sequences using 10x Cell Ranger (v.2.1). Mapping to the human-chimp-macaque consensus genome and generation of count matrices was then performed using the same Cell Ranger, with the GENCODE v.27 human annotation provided. Nuclei were assigned to species based on species-specific sites using a two-step approach by separating all great ape from macaque nuclei first and subsequently assigning nuclei to either human or chimp/bonobo. Nuclei with a support of less than 80% for either of the groups were removed from further analysis. Moreover, nuclei with fewer than 200 and more than 6,000 genes detected, as well as those with more than 5% detected transcripts being transcribed from mitochondria, were removed from further analyses.

The full snRNA-seq dataset including all species was further analysed using Seurat (v.3) (Supplementary Table 13). Single-nucleus expression values were normalized and highly variable genes were identified using a variance stabilizing function to detect the top 2,000 variable genes (Supplementary Table 12). Data were then integrated by finding corresponding anchors between the species using 30 dimensions. Scaling and PCA were performed using the integrated data. The top-20 PCs were used to identify neighbours of cells and clusters and to visualize the clustering using *t*-SNE embedding. Cluster identities were assigned using unbiased identification using cluster markers by running Seurat's FindAllMarkers function (Wilcoxon test, min.pct = 0.25, min logFC = 0.25) using non-integrated expression values, known marker genes reported elsewhere[66,67] and by cell-type prediction using Seurat's TransferData function to anchor to the published drop-seq based human adult frontal cortex snRNA-seq data[67]. Two potential doublet clusters (c11, c19) were excluded from further analysis. For analysis of the major cell classes (excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, microglia, endothelial cells) subtype clusters were combined and cell-type markers recalculated using Seurat's FindAllMarkers function (Wilcoxon test, min.pct = 0.25, min logFC = 0.25) using non-integrated expression values (Supplementary Table 14).

As nuclei of the three species have significantly different transcriptome coverage, pseudo-nuclei were constructed for more robust transcriptome measurement, as well as for more fair and efficient comparison, using a similar procedure as described above to generate pseudocells, under the constraint of merging only nuclei from the same segment of the same sample and grouped in the same cell cluster. The probabilities of nuclei selected as pseudo-nuclei seed were 1/13 for human, 1/8 for chimpanzee and 1/10 for macaque.

Reads of the bulk RNA-seq samples were mapped to the human-chimpanzee-macaque consensus genome using STAR (v.2.6.1d). The Python utility hiseq-count was used to count the numbers of uniquely mapped reads of genes annotated in GENCODE v.27 human annotations. DESeq2 was used for normalization and retrieving FPKM as the expression level measurement.

To determine the laminar origin of each segment, genes with segment-dependent expression were first screened for each cortical cube. In brief, an ANCOVA analysis was applied to compare two models: the natural spline (df = 6) linear model with $\log_{10}$-transformed FPKM as the response and the segment order as the variable; the null model of expression values without any relationship with segments. For each of the resulted gene, its enriched segments in the cube were identified as the segments with the gene's expression at least one standard deviation higher than the mean across segments. Genes with enriched expression at each segment were then overlapped with the layer markers identified previously[3]. Segments with enriched genes significantly overlapping with markers of only one layer were seen as pure-layer original, others were seen as mixture of multiple layers. For each mixture segment, a quadratic-programming-based transcriptome deconvolution[68] was applied to determine the relative contribution of the enriched layers. A layer index was then obtained for each segment, as the average layers weighted by contributed proportion of the enriched layers.

**Estimation of cell-type distribution across cortical layers and gene-expression patterns in neurons across cortical layers.** To estimate the cell-type composition of each layer, nuclei from each sample were randomly assigned to one layer, based on the layer mixture proportions estimated above. The proportion of each of the six major cell classes: excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocyte precursor cells (OPCs), microglia and endothelial cells, was then calculated for nuclei assigned to each layer in human. This procedure was repeated 100 times, with the resulting average as the final estimation. The laminar distribution of each cell cluster was also estimated based on the described procedure. In addition, a subsampling procedure with replaceable manner of the same number of nuclei ($n = 200$) from each layer was further applied to each of the 100 nuclei layer random assignment to control differences on the detected nuclei number of each layer.

To get more precise estimation of layer origins on the nuclei level for excitatory and inhibitory neurons, both of which show a distinct layer distribution pattern across different subtypes, we trained an elastic net linear regression model ($\alpha = 0.5$) on excitatory and inhibitory neurons separately, with the sample layer indices as the training response and expression levels of the highly variable genes as the variables. To enhance model robustness, pseudo-nuclei from all the three species together were used for model trainings. The trained models were then applied to the excitatory and inhibitory pseudo-nuclei again. The predicted layer indices were used as the estimated relative laminar location of the pseudo-nuclei. The projection of the predicted layer indices to layers were done by averaging expression patterns of markers of different layers[3].

**Differential expression analysis between human and chimpanzee cell types in adult brains and determination of their species-specificity.** Owing to the sparse nature of the snRNA-seq data and the unequal coverage of nuclei from different species, commonly used statistical tests for differential expression analysis (for example, Wilcoxon's rank-sum test) failed to provide reliable estimation of DE, even with the state-of-art VST normalization methods[69]. As detection rates of genes are correlated with their expression levels[69], we therefore compared gene expression levels of the same cell type in human and chimpanzee by comparing their detection rates, using a GLM-ANCOVA analysis similar to the one described above to identify genomic regions with differential accessibility. In brief, the pseudo-nuclei expression matrix was binarized. A binomial GLM model was trained for each gene, with its detection as the response variable and species of pseudo-nuclei as the independent variable. This model was compared to the null model with the species variable replaced by a scaling coefficient. The scaling coefficient is fixed to one for human pseudo-nuclei and $p_c/p_h$ for chimpanzee pseudo-nuclei, where $p_c$ and $p_h$ are the average detected gene numbers across pseudo-nuclei involved in the test in chimpanzee and human, respectively.

While the described DE test was applied to four cell classes with sufficient numbers of pseudo-nuclei: excitatory neurons, inhibitory neurons, astrocytes and oligodendrocytes, the heterogeneity within the two neuron types, as well as their uneven distributions in human and chimpanzee, needed to be considered. A subsampling procedure with replaceable manner was therefore applied. In every subsampling, an equal number of pseudo-nuclei ($n = 200$) from each species were sampled, with pseudonuclei in clusters annotated as the cell class of interest sharing equal probability being selected. The described DE test was then applied to the sampled nuclei of this cell class. This subsampling procedure was repeated 100 times, and DE genes of each cell class were defined as genes with significant DE (BH-corrected $P < 0.005$) in at least 80 times of the subsampling. Additional filtering was then applied, requiring the same direction of human–chimpanzee difference on detection rates and VST-normalized expression values.

Macaque pseudo-nuclei were then introduced to investigate species specificity of the identified DE. A similar procedure sampling the same number of pseudo-nuclei from clusters annotated to be the same cell class was repeated 100 times to the macaque pseudo-nuclei. For each sampling, average VST-normalized expression values were calculated for each cell class in human, chimpanzee and macaque, with which differences between human and macaque ($d_{\mathrm{HM}}$), as well as between chimpanzee and macaque ($d_{\mathrm{CM}}$), were calculated. The identified human–chimpanzee DE was defined as human-specific if $|d_{\mathrm{HM}}| > 4 \times |d_{HC}|$. Genes with chimpanzee-specific DE were identified in the same way (Supplementary Table 15).

**Statistics and reproducibility.** In the box plots of Figs. 1g, 2d, boxes represent IQR and whiskers represent minimum and maximum with outliers removed. In the box plot of Fig. 4c, boxes represent IQR and whiskers represent 1.5 × IQR. In the bean plots of Extended Data Fig. 3g, shapes of beans represent Pearson correlation distributions and dashed lines represent medians of groups.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Sequence data that support the findings of this study have been deposited in ArrayExpress with the accession codes E-MTAB-7552 (single-cell RNA-seq data based on 10x Genomics), E-MTAB-8234 (single-cell RNA-seq data based on Fluidigm C1/Smart-seq2), E-MTAB-8089 (single-cell ATAC-seq of human organoids), E-MTAB-8043 (single-cell ATAC-seq of chimpanzee organoids), E-MTAB-8083 (single-cell ATAC-seq of bonobo organoids), E-MTAB-8087 (single-cell ATAC-seq of macaque organoids), E-MTAB-8228 (the bulk ATAC-seq data), E-MTAB-8230 (snRNA-seq data of the adult brain samples) and E-MTAB-8231 (bulk RNA-seq data). The expression data are also available for exploration in scApeX via the link https://bioinf.eva.mpg.de/shiny/sample-apps/scApeX/.

## Code availability

The computational code used in this study is available at GitHub (https://github.com/quadbiolab/primate_cerebral_organoids) or upon request.

30. Kilpinen, H. et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017).
31. Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
32. Okita, K. et al. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells* **31**, 458–466 (2013).
33. Lancaster, M. A. & Knoblich, J. A. Generation of cerebral organoids from human pluripotent stem cells. *Nat. Protoc.* **9**, 2329–2340 (2014).
34. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
35. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).
36. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
37. Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
38. He, Z., Bammann, H., Han, D., Xie, G. & Khaitovich, P. Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. *RNA* **20**, 1103–1111 (2014).
39. Dougherty, M. L. et al. Transcriptional fates of human-specific segmental duplications in brain. *Genome Res.* **28**, 1566–1576 (2018).
40. Fiddes, I. T. et al. Human-specific *NOTCH2NL* genes affect Notch signaling and cortical neurogenesis. *Cell.* **173**, 1356–1369 (2018).
41. Dennis, M. Y. et al. Evolution of human-specific neural *SRGAP2* genes by incomplete segmental duplication. *Cell* **149**, 912–922 (2012).
42. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
43. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
44. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
45. Renaud, G., Stenzel, U., Maricic, T., Wiebe, V. & Kelso, J. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770–772 (2015).
46. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
48. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
49. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
50. Pliner, H. A. et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871 (2018).
51. Preissl, S. et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat. Neurosci.* **21**, 432–439 (2018).
52. Suzuki, K. et al. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell Stem Cell* **15**, 31–36 (2014).
53. Mayer, C. et al. Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
54. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
55. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
56. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
57. Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).
58. Gittelman, R. M. et al. Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res.* **25**, 1245–1255 (2015).
59. Cagan, A. et al. Natural selection in the great apes. *Mol. Biol. Evol.* **33**, 3268–3283 (2016).
60. Peyrégne, S., Boyle, M. J., Dannemann, M. & Prüfer, K. Detecting ancient positive selection in humans using extended lineage sorting. *Genome Res.* **27**, 1563–1572 (2017).
61. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
62. Chintalapati, M., Dannemann, M. & Prüfer, K. Using the Neandertal genome to study the evolution of small insertions and deletions in modern humans. *BMC Evol. Biol.* **17**, 179 (2017).
63. McLean, C. Y. et al. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011).
64. Fu, Y. et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.* **15**, 480 (2014).
65. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).

66. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
67. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
68. Yu, Q. & He, Z. Comprehensive investigation of temporal and autism-associated cell type composition-dependent and independent gene expression changes in human brains. *Sci. Rep.* **7**, 4121 (2017).
69. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Preprint at *bioRxiv* https://doi.org/10.1101/576827 (2019).
70. Sloan, S.A. et al. Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron* **95**, 779–790 (2017).

**Acknowledgements** We thank D. Wollny, A. Brazovskaya, K. Köhler, T. Schaffer, B. Schellbach, A. Weihmann, R. Schultz, I. Bünger, M. Dannemann, R. Snabel, B. Vernot, W. Hevers, M. Schörnig, J. Kelso and K. Sekine for their help with this project; A. Fischer, M. Halbwax, K. Köhler and the Tchimpounga Sanctuary for support with generating the JoC iPSC line; L. Berninger and J. Peters for contributing Sc102a1 and SandraA organoids; B. B. Lake for sharing nuclei isolation protocols; and A. Fischer, M. Halbwax and the Lola ya Bonobo Sanctuary in Congo for contributing bonobo tissue for single-nucleus experiments. Karyotyping was supported by the Stem Cell Engineering Facility of CMCB at Technische Universität Dresden. Sorting was in part performed at the CUDZ at the Veterinary Medicine Faculty at the University of Leipzig. The silhouette images in the figures are credited to istockphoto.com for human (cole matt) and chimpanzee (A-digit), and flaticon.com for macaque (Freepik). This work was supported by the Max Planck Society (B.T.), Chan–Zuckerberg Initiative (B.T. and J.G.C.), European Research Council (Anthropoid, J.G.C.; Organomics; B.T.) and the NOMIS Foundation (S.P.). S.K. was supported by the Boehringer Ingelheim Fonds.

**Author contributions** S.K. and M.J.B. grew organoids with assistance from A.W., L.S. and M.H. S.K. performed scRNA-seq and snRNA-seq with assistance from M.S. M.J.B. performed scATAC-seq. Z.H., M.J.B. and S.K. analysed the data. F.S.C. and M.H. performed immunohistochemical stainings. J.S.F. compared organoid scRNA-seq data to mouse voxel maps. P.G. dissected and sliced tissue for snRNA-seq. D.H. and Z.Q. performed bulk RNA-seq of adult tissue. S.K., M.J.B., Z.H., B.T. and J.G.C. designed the study and wrote the manuscript with support from P.K., W.B.H. and S.P.

**Competing interests** The authors declare no competing interests.

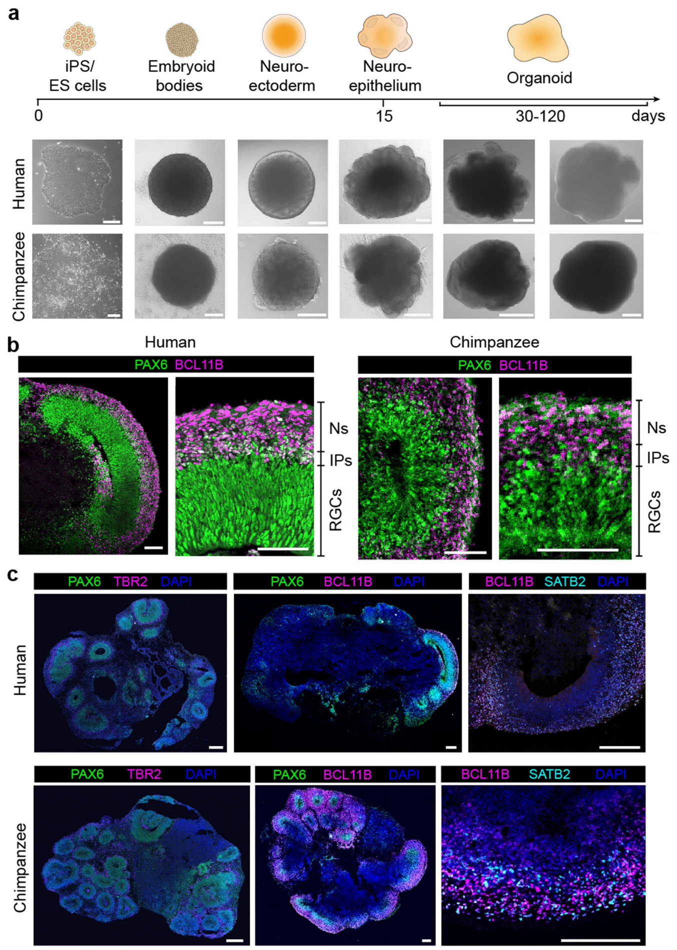
**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1654-9.

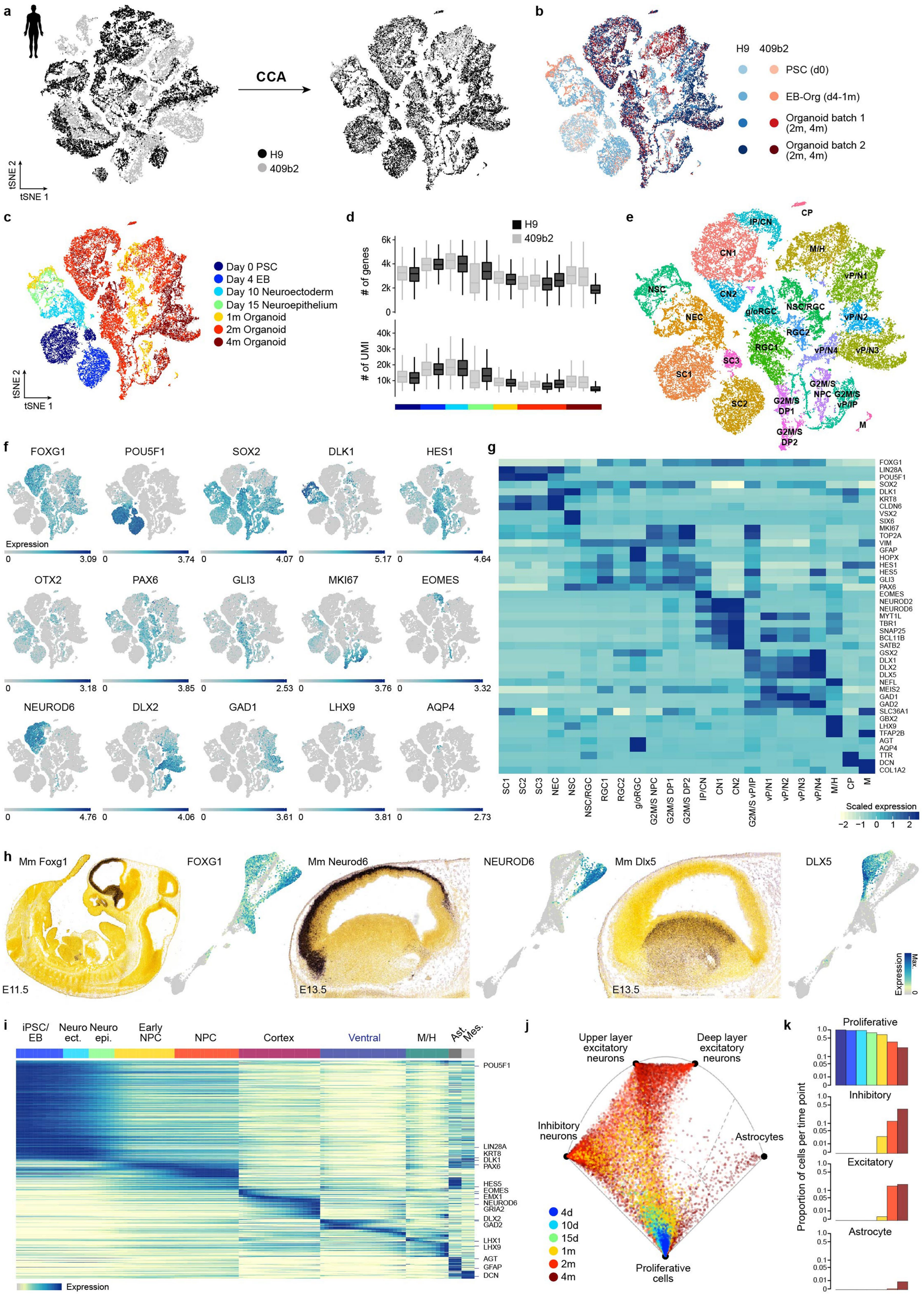
**Correspondence and requests for materials** should be addressed to Z.H., B.T. or J.G.C.

**Peer review information** *Nature* thanks Christopher Walsh and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints.

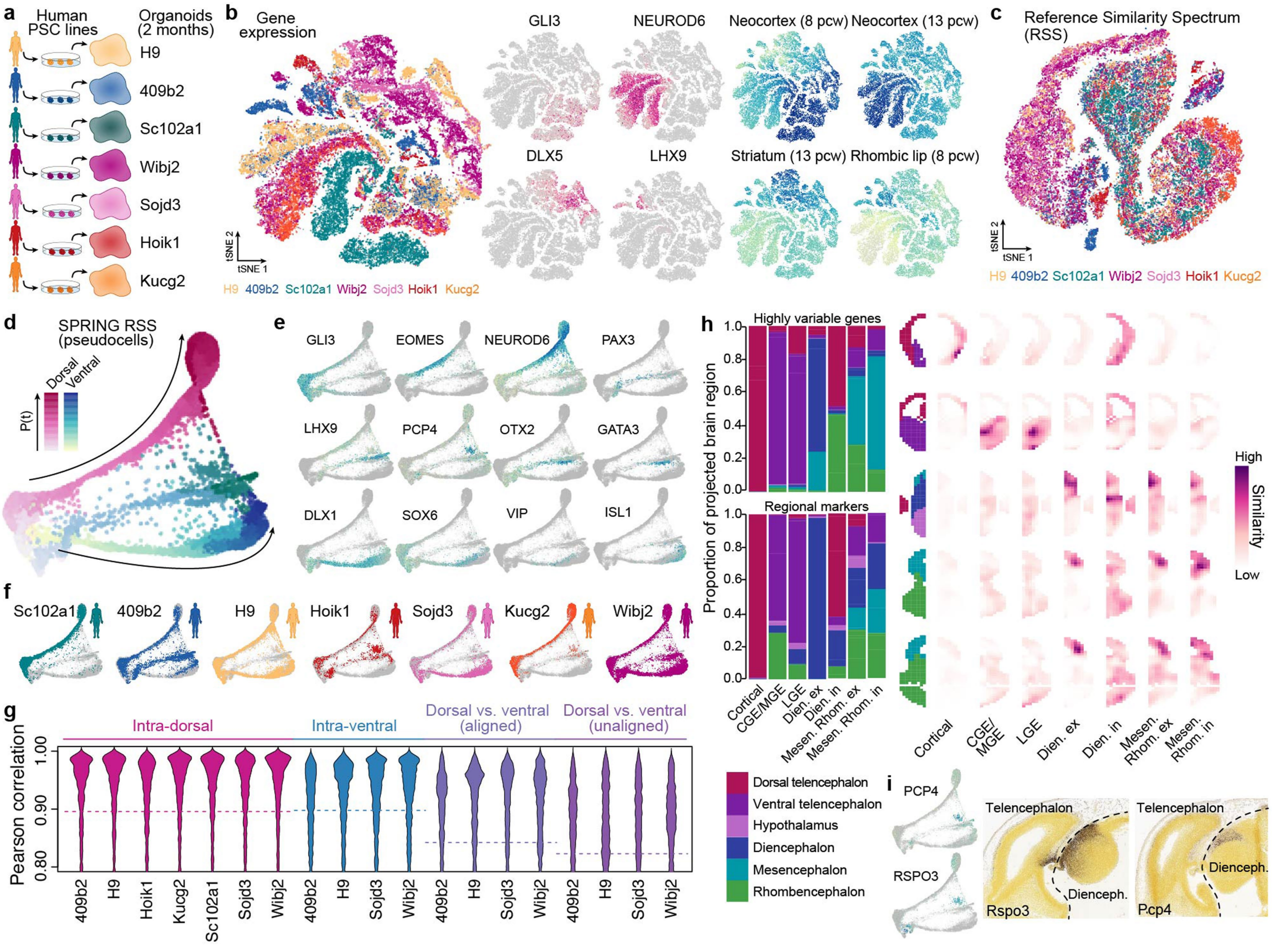


**Extended Data Fig. 1 | Differentiation and immunohistochemical characterization of human and chimpanzee cerebral organoids. a**, Phase contrast (PSC to neuroepithelium; scale bars, 200 μm; H9 for human, SandraA for chimpanzee) and bright-field images (organoid; scale bars, 1 mm; H9 and Wibj2 for human, JoC and SandraA for chimpanzee) showing examples of different stages of organoid development for human and chimpanzee. **b**, Immunohistochemical staining for *PAX6* (green) and *BCL11B* (also known as *CTIP2*) (pink) of a 63-day human organoid from iPSC line 409b2 (left) and a 63-day chimpanzee organoid from iPSC line SandraA (right), with a magnified view into a cortical-like region (scale bars, 100 μm). **c**, Immunohistochemical staining of human (top left, Sc102a1, 50-day; top middle, 409b2, 63-day, the same organoid as the human organoid in **b**) and chimpanzee (bottom left, SandraA, 50-day; bottom middle, SandraA, 63-day) organoids (scale bars, 200 μm) for progenitor (*PAX6*) intermediate progenitor (*TBR2*) and deep-layer neurons (*BCL11B*) in whole organoids. Staining for deep-layer (*BCL11B*) and upper-layer (*SATB2*) neuron markers for human (top right, Sc102a1, 45-day) and chimpanzee (bottom right, SandraA, 63-day) organoids.

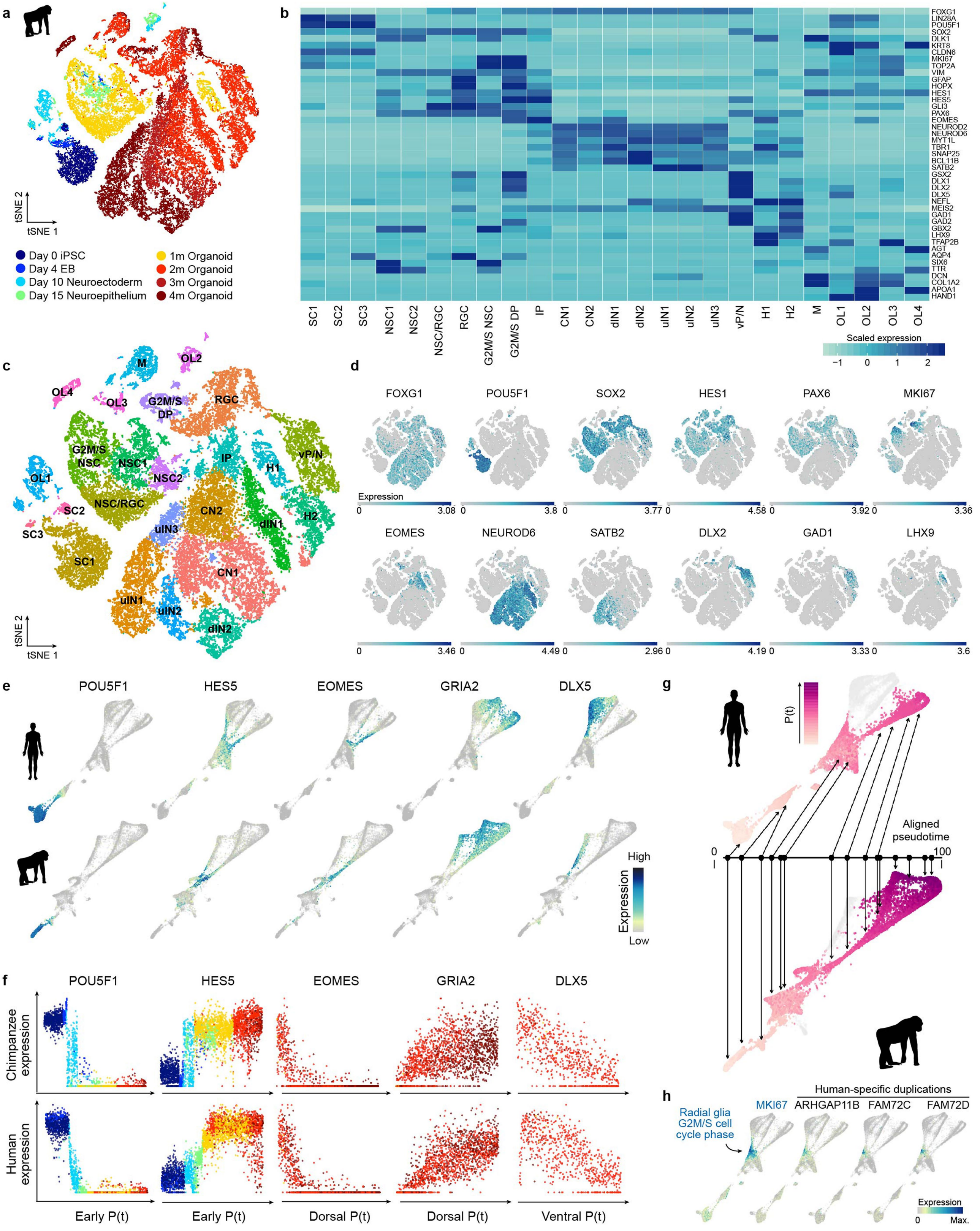


**Extended Data Fig. 2** | See next page for caption.

**Extended Data Fig. 2 | Heterogeneity analysis during human cerebral organoid development from pluripotency. a**, Cells from different human cell lines (23,226 cells from H9 and 20,272 cells from 409b2) were integrated using CCA and visualized using *t*-SNE. **b**, *t*-SNE coloured on the basis of cell line and batch. **c**, *t*-SNE coloured on the basis of time point. Heterogeneity analysis was performed on combined cells from day 0 of differentiation to 4-month-old organoids for iPSC and ESC-derived cells. **d**, Distribution of number of genes and UMIs for different time points and cell lines. **e**, Clustering was performed using the top-20 PCs as input for *t*-SNE and cluster names were assigned on the basis of expression of cluster marker genes and known marker genes. SC, stem cells; NEC, neuroectoderm-like cells; NSC, neural stem cells; g/oRGC, gliogenic/outer radial glia cells; G2M/S NPC, neural progenitor cells in G2M/S phase; G2M/S DP, dorsal progenitor cells in G2M/S phase; CN, cortical neurons; G2M/S vP, ventral progenitors in G2M/S phase; M/H, midbrain/hindbrain; CP, choroid plexus; M, mesenchymal-like cells. **f**, *t*-SNE plots coloured by expression level of selected marker genes on the basis of non-integrated expression values. **g**, Heat map showing averaged cluster expression for representative marker genes for clusters ordered according to their differentiation time from early to later stages and regional identity from dorsal to ventral forebrain and non-forebrain cells. **h**, In situ hybridization images from the Allen Developing Mouse Brain Atlas (available from https://developingmouse.brain-map.org/) showing expression of *Foxg1*, *Neurod6* and *Dlx5* in the mouse developing forebrain and human whole-trajectory SPRING plots coloured by the corresponding genes. **i**, Pseudotemporal expression of example genes marking different stages of development over the whole human cerebral organoid developmental trajectory. **j**, Umbrella plot showing the similarity of each organoid cell to a cell 'prototype' generated from a reference scRNA-seq cell atlas of the human fetal cortex[19]. **k**, Plots show the proportion of organoid cells per time point that match a reference prototype.

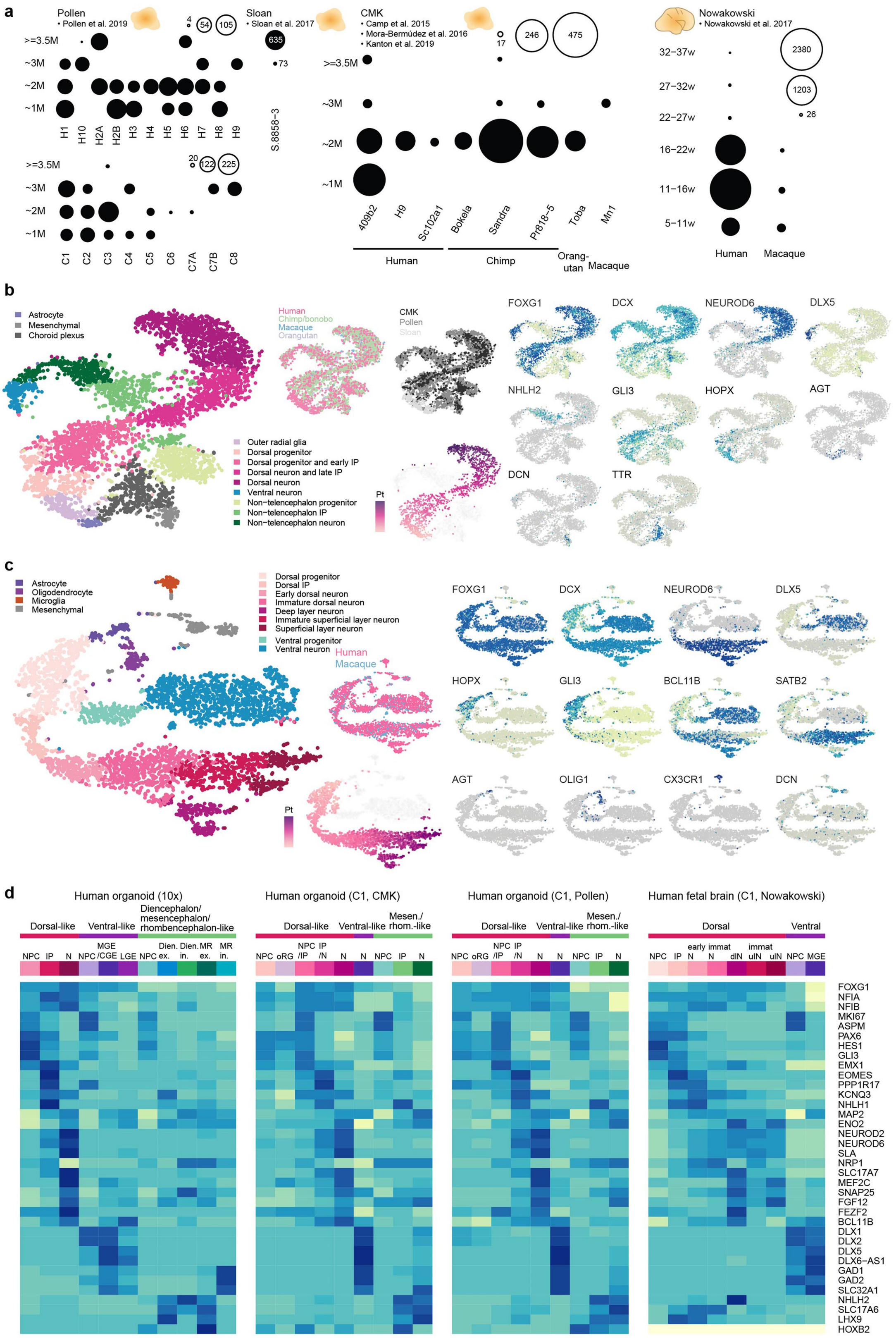


**Extended Data Fig. 3 | Analysis of human cerebral organoid single-cell transcriptomes from seven individuals. a**, scRNA-seq was performed on two-month-old cerebral organoids from one human ESC and six iPSC lines. **b**, All data (49,153 cells) were combined and cell heterogeneity was assessed using *t*-SNE with the top 20 PCs as the input. Cells are also coloured by marker gene expression and RSS. **c**, *t*-SNE plot with RSS against Brainspan fetal reference data as the input (RSS-*t*-SNE), coloured by cell lines. Cells from different lines are well integrated. **d**, SPRING plot of two-month-old human organoid pseudocells (9,650), coloured by neuronal trajectory branches and pseudotimes. **e**, SPRING plots of two-month-old human organoid cells, coloured by marker gene expression. **f**, SPRING plots coloured by cell line show contributions of each line to different branches of the trajectory. **g**, Correlations of expression trajectories of genes with pseudotime-dependent expression patterns between cortical cells from each line to the others (pink), ventral cells from each line to others (blue), and cortical and ventral cells from the same lines after or before aligning the cortical and ventral pseudotimes (purple). **h**, Spatial location inference of neuron subtypes in human cerebral organoids. Left, bar plots show proportion of cells of each cell type that show highest gene-expression-pattern similarity to the average expression patterns in different structures, on the basis of the processed in situ hybridization image data (E13.5) provided in the Developing Mouse Brain database of Allen Brain Atlas (available from https://developingmouse.brain-map.org/). Expression similarity was calculated based on highly variable genes of the scRNA-seq data (top) or regional markers defined with the in situ hybridization data (bottom). Right, correlation patterns of average regional marker gene expression of each neuron subtype to voxels in five example sections (E13.5), as well as the structural annotation of the sections. **i**, Expression of two marker genes of diencephalon inhibitory neurons (*PCP4* and *RSPO3*) in the SPRING embeddings, and their spatial expression patterns in E13.5 mouse brain (data from Allen Brain Atlas, available from https://developingmouse.brain-map.org/).



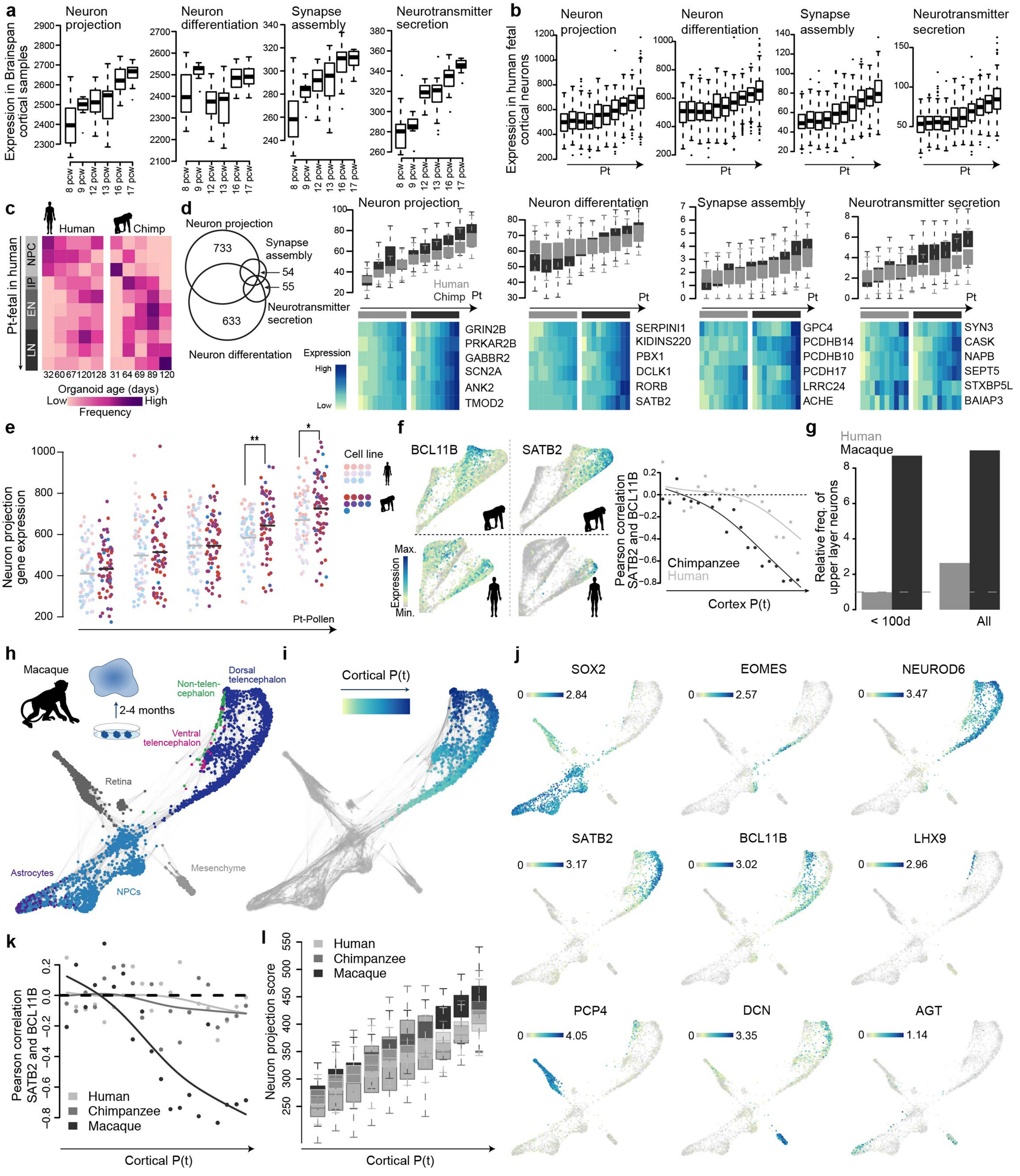
**Extended Data Fig. 4** | See next page for caption.

**Extended Data Fig. 4 | Heterogeneity analysis during chimp cerebral organoid development from pluripotency.** **a**, Heterogeneity analysis for iPSC-derived chimpanzee cells (36,884) from day 0 of differentiation to 4 months of organoid development for one cell line (SandraA). **b**, Heat map visualizing averaged cluster expression for marker genes with columns ordered based on differentiation progress from early-to-late time points and regional identity sorted from dorsal to ventral forebrain to non-forebrain cells and non-ectodermal-derived cells. **c**, Cluster identification and *t*-SNE using the top-15 PCs for clustering. Cluster assignment was on the basis of cluster markers as well as expression patterns of known marker genes. SC, stem cells; G2M/S DP, dorsal progenitors in G2M/S phase; dlN, deep-layer neuron; ulN, upper-layer neurons; vP/N, ventral progenitor/neuron; M – mesenchymal-like cells; OL, off-lineage cells. **d**, *t*-SNE plots coloured on the basis of gene expression of representative marker genes used to assign cluster identities. **e**, SPRING plots of whole developmental trajectory for human and chimpanzee coloured by marker genes. **f**, Pseudotemporal gene-expression patterns showing marker genes for early, dorsal and ventral branches for human and chimpanzee. **g**, Schematic showing alignment of human and chimpanzee pseudotimes after combining pseudocells from the early stages and the dorsal forebrain lineage. The later chimpanzee pseudotime points fail to align with human pseudocells. **h**, Expression of genes duplicated in humans projected onto the human whole-lineage SPRING analysis, with the G2M/S phase marker MKI67 shown as a reference.



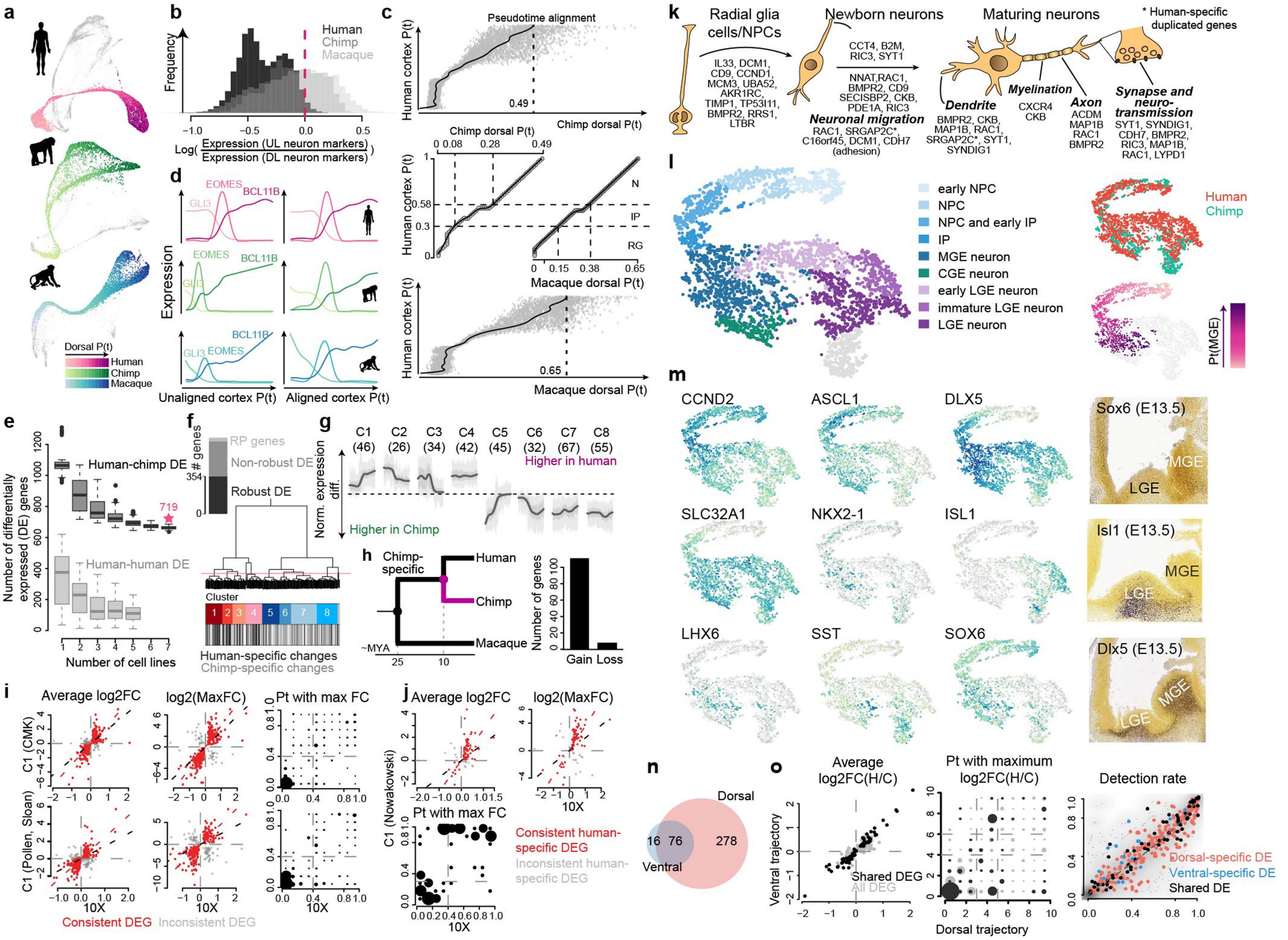
**Extended Data Fig. 5** | See next page for caption.

**Extended Data Fig. 5 | Analysis of cell-type heterogeneity of cerebral organoids and fetal cortical tissues based on scRNA-seq data from Fluidigm C1. a**, Overview of the Fluidigm C1 scRNA-seq data. Each dot represents a cerebral organoid or fetal brain sample from one cell line or species at a certain age, with its size showing the number of cells measured. The left panel shows organoid sample information as published in Pollen et al. (2019)[16] (excluding redundant cells from Camp et al. (2015)[11] and Mora-Bermudez et al. (2016)[15]), including the data initially published in Sloan et al. (2017)[70]. The middle panel shows organoid sample information generated in Camp et al. (2015)[11], Mora-Bermudez et al. (2016)[15] and in this study. The right panel shows fetal prefrontal cortex sample information reported in Nowakowski et al. (2017)[19]. **b**, All cerebral organoid data (5,838 cells) were combined and cell heterogeneity was assessed using *t*-SNE with the RSS profiles to the fetal Brainspan data as the input. Cells are coloured by cell type or cluster, species, institutions generating the data, dorsal trajectory pseudotimes and marker gene expression. **c**, *t*-SNE plots for all fetal brain data (5,080 cells) to assess cell heterogeneity, with the RSS profiles to the fetal Brainspan references as the input. Cells are coloured by cell type or cluster, species, dorsal excitatory neuron trajectory pseudotimes and marker gene expression. **d**, Heat map showing marker gene-expression patterns across different cell types in the droplet-based organoid scRNA-seq data generated in this manuscript and the above described C1-based scRNA-seq data.



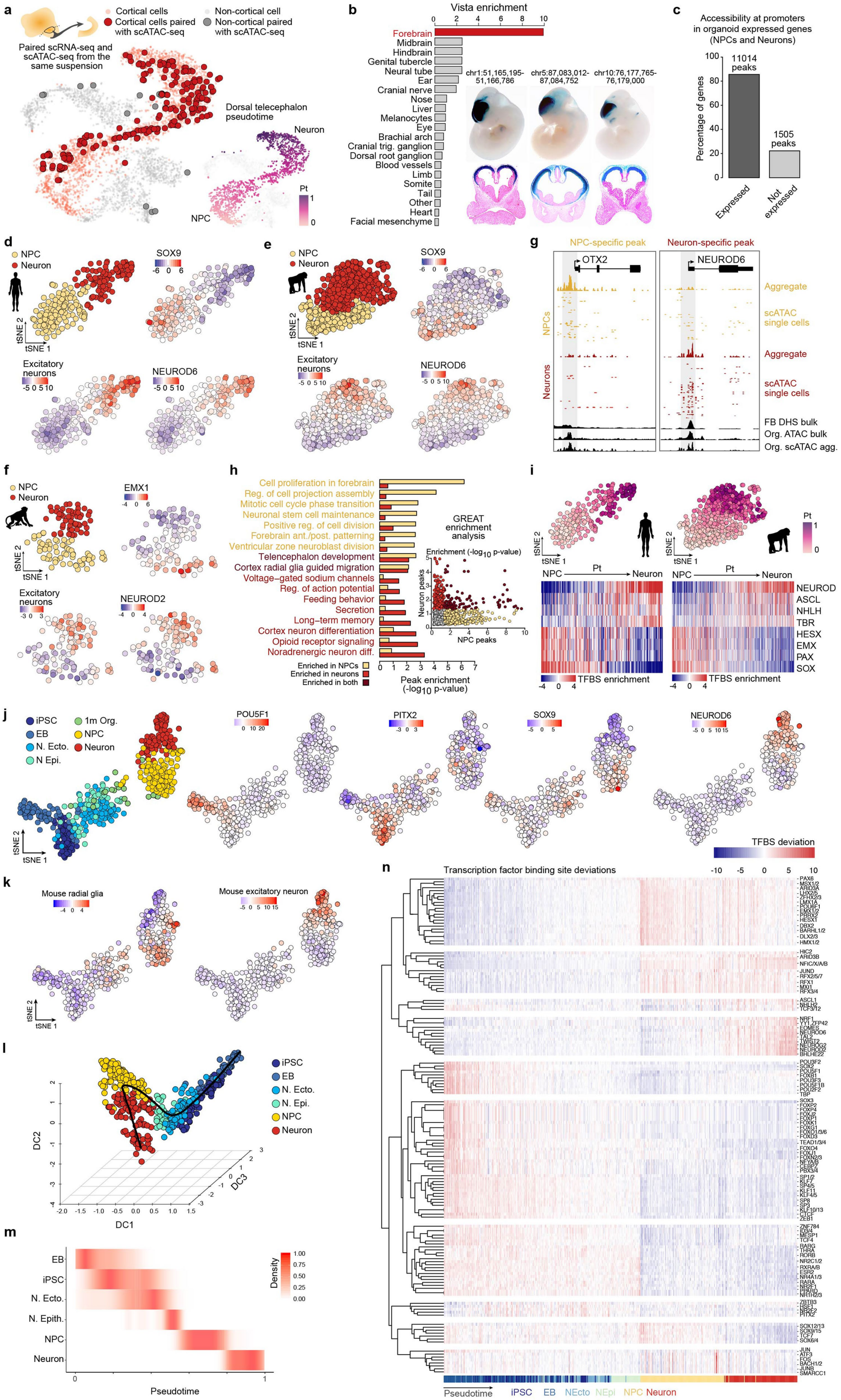
**Extended Data Fig. 6** | See next page for caption.

**Extended Data Fig. 6 | Analysis of neuron maturation timing difference in human, chimpanzee and macaque cerebral organoids.** **a**, Box plots (IQR with minimum and maximum, outliers removed) showing sum expression levels (in reads per kilobase of transcript per million (RPKM)) of genes with GO annotation neuron projection (1,487 genes), neuron differentiation (1,367 genes), synapse assembly (168 genes) and neurotransmitter secretion (169 genes) in bulk RNA-seq data from Brainspan fetal cortical samples from 8 PCW to 17 PCW. **b**, Box plots showing sum expression levels of the same gene lists in fetal human dorsal excitatory neurons along the estimated developmental pseudotimes (Nowakowski et al. (2017) dataset[19]). **c**, Projection of human and chimpanzee organoid cells to human fetal brain data reveals higher similarity of chimpanzee organoid cells to later stages of development compared to human organoid cells. **d**, Box plots showing sum expression levels of genes with specific annotation to only one of the four GO terms in human and chimpanzee pseudocells (1,791 human and 4,304 chimp) along the cortical pseudotimes. Heat maps showing expression of example genes from the GO terms for human and chimp along pseudotime bins. The Venn diagram on the left shows the overlap of genes related to the four GO terms. **e**, Distribution of neuron projection scores of human and chimpanzee cortical cells (388 human and 355 chimp) reported in Pollen et al. (2019) along the cortical pseudotimes[16]. Each dot represents one cell, and is coloured by the organoid cell line. Light colours represent human cell lines and dark colours represent chimpanzee ones. Two-sided Wilcoxon's rank-sum test ($*P = 0.013$ and $**P = 0.004$). **f**, Observed timing difference of upper-deeper layer specification in human and chimpanzee cerebral organoids from 10x Genomics data generated in this study. The left panel shows expression of cortical deep (*BCL11B*, left) and upper (*SATB2*, right) layer marker genes projected onto the chimpanzee (top) and human (bottom) SPRING plot. *BCL11B* and *SATB2* become anti-correlated in their pseudotemporal expression profile in both human and chimpanzee (right), while the onset of anti-correlation happens earlier in chimpanzee than in human. **g**, Abundance of upper-layer neurons relative to deeper-layer neurons in human and macaque fetal prefrontal cortex samples[19] in Nowakowski et al. (2017) grouped by early time points ($<$100 days old) or all time points combined. **h**, scRNA-seq was performed on two-to-four-month cerebral organoids from a macaque iPSC line. The SPRING plot of pseudocells (2,913) was constructed with the top-20 PCs as the input. The heterogeneity analysis suggests multiple cell types in the macaque organoids, including cortical neurons, NPCs, astrocytes and other cell types such as retina and mesenchyme-like cells. **i**, SPRING plot coloured by pseudotimes of cortical pseudocells, which are the pseudocells' quantiles of diffusion component (DC) 1 of the cortical pseudocells diffusion map. **j**, SPRING plot coloured by marker gene expression. **k**, The onset of anti-correlation between *SATB2* and *BCL11B* occurs earlier along the macaque pseudotime (1,107 pseudocells), relative to human (1,118 pseudocells) and chimpanzee (1,645 pseudocells), when focusing on the two-month cerebral organoids. **l**, Box plots (box shows IQR and whiskers show 1.5 × IQR) showing the neuron projection scores in human, chimpanzee and macaque along the unaligned cortical pseudotimes.



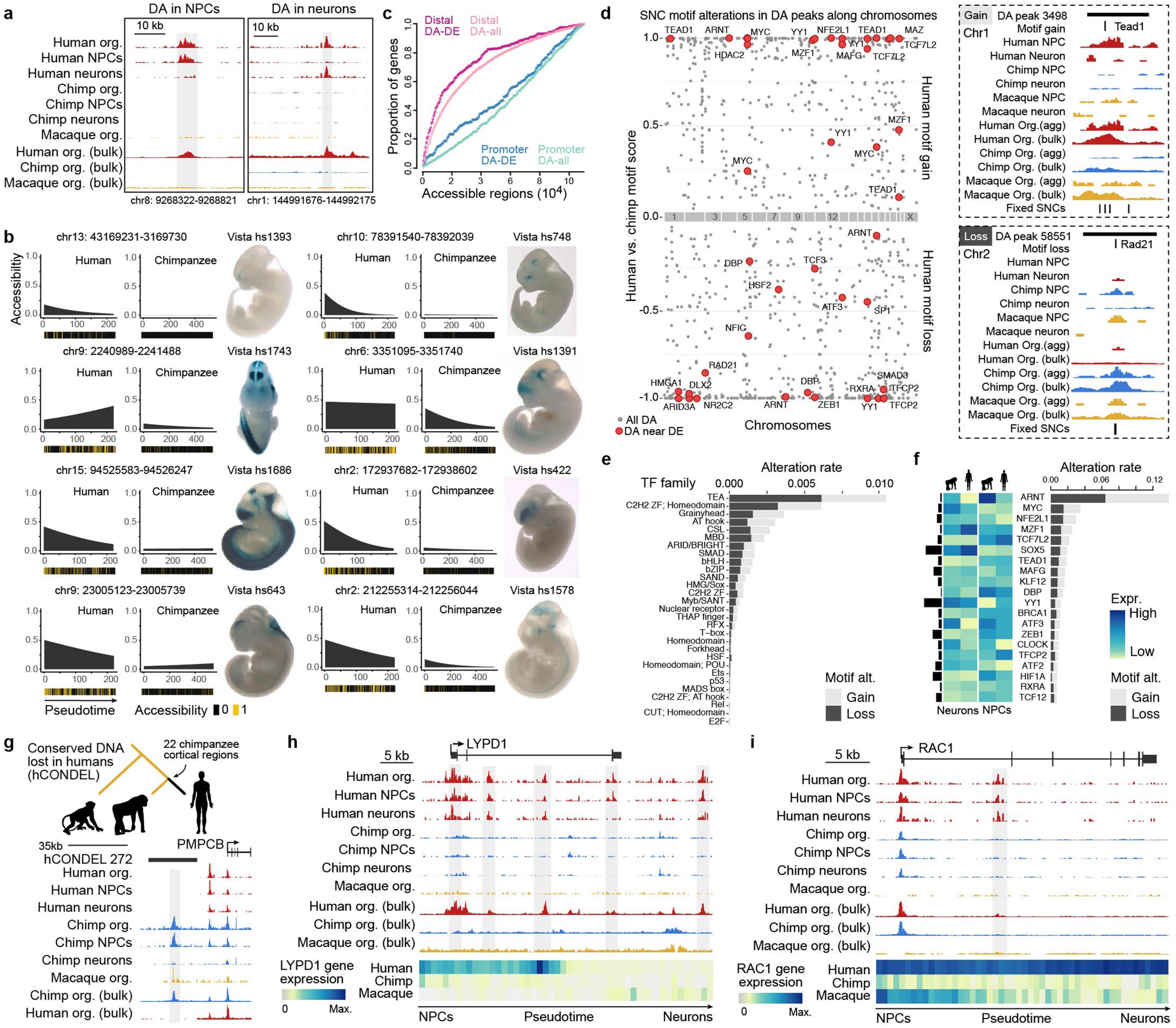
**Extended Data Fig. 7** | See next page for caption.

**Extended Data Fig. 7 | Pseudotime alignment between primates and differential expression between human and chimpanzee. a**, SPRING plots visualizing the *k*NN networks of human (10,063) and chimpanzee (5,612) pseudocells, and macaque cells (6,580), which represent NPCs and neurons of different brain regions. Cortical NPCs and neurons are coloured by their pseudotimes. **b**, Ratios of upper layer (UL) to deeper layer (DL) neuron marker expression in human (black), chimpanzee (dark grey) and macaque (light grey) organoids. The dashed line indicates the cut-off applied to human pseudocells to filter out those representing UL neurons. **c**, Truncated dynamic time warping (DTW)-based alignment was applied to align human, chimpanzee and macaque cortical pseudotime courses. Two support vector regression models were trained to predict chimpanzee (top) and macaque (bottom) pseudotimes of human pseudocells. A constrained B-splines regression model was fitted to determine the trimming point at the chimpanzee and macaque pseudotime courses, respectively. An end-to-end DTW-based alignment was applied to the human pseudotime course to the trimmed chimpanzee and macaque pseudotime courses for the final alignments (middle). **d**, Pseudotemporal expression profiles of *GLI3*, *EOMES* and *BCL11B* along the human, chimpanzee and macaque cortical pseudotimes, before (left) and after (right) the pseudotime alignment procedures. **e**, Robustness and false-positive rate of differential pseudotemporal expression between human and chimpanzee based on the number of cell lines involved in the analysis with constrained replaceable pseudocell subsampling. In each subsampling, pseudocells representing cells from a certain number of human lines were sampled in a replaceable manner to recapitulate pseudocell distribution along pseudotime course of the chimpanzee pseudocells. Differential expression analysis was applied to compare all chimpanzee pseudocells and the sampled human pseudocells to estimate robustness to cell line numbers (dark grey boxes), and to compare sampled human pseudocells to human pseudocells from another two lines sampled with the same procedure to estimate false-positive rate (light grey boxes). In box plots, boxes represent 100 times of subsampling IQR, the line represents 1.5 × IQR and dots represent outliers. **f**, Robustly detected human–chimpanzee differentially expressed genes (robust DE genes) are defined as the non-ribosomal genes which were detected as DE in at least 80% of the subsampling-based human–chimpanzee DE analysis using any number of human lines (black). The dendrogram shows the hierarchical clustering of robust DE genes, based on their human–chimpanzee pseudotemporal DE patterns along the aligned pseudotimes of cortical organoid pseudocells, resulting in eight clusters of robust DE genes. **g**, Pseudotemporal differential expression patterns between human and chimpanzee (without including macaque cells) of the eight clusters of genes along the pseudotimes of cortical organoid pseudocells with 50% and 95% confidence intervals shown in dark and light grey, respectively. Numbers of genes in each cluster are shown in parenthesis. **h**, Number of differentially expressed genes in chimpanzee versus human and macaque comparison grouped by gain or loss of expression in chimpanzees. A gain of expression specifically in chimpanzees is more likely than a loss of expression pattern conserved in the other primates. **i**, Comparison of the reported human–chimpanzee pseudotemporal differential expression based on 10x Genomics data with the Fluidigm C1-based scRNA-seq data of human and chimpanzee cerebral organoids. The two rows show the results based on C1 data generated in this manuscript and combined with data from refs. [11,15,16]. The first two columns show estimated human–chimpanzee differential expression directionality and magnitude in the reported droplet-based scRNA-seq data and the C1-based measurement, with the first column presenting the generalized differential expression along the whole cortical pseudotimes, and the second column presenting the maximum differential expression along the pseudotimes. The red dots represent consistently differentially expressed genes, which have consistent differential expression directionalities in the two datasets. The right panel shows pseudotime intervals with the largest human–chimpanzee differential expression in the two datasets in comparison to the consistent differentially expressed genes. Dot sizes represent frequencies. **j**, Comparison of the estimated human–macaque differential expression directionality and magnitude of the human-specific differentially expressed genes using human and macaque fetal prefrontal cortex scRNA-seq data[16,19]. **k**, Functional annotations of genes with human-specific expression patterns based on GO annotations related to brain development and neurogenesis. Only the human-specific differentially expressed genes with consistent human–chimpanzee or human–macaque differential expression detected in at least one of the three C1-based scRNA-seq datasets are shown. **l**, Ventral telencephalon cell heterogeneity in organoids was investigated by *t*-SNE embeddings with RSS profiles of human (3,385) and chimpanzee ventral (773) pseudocells combined as the input. Pseudocell clusters were annotated on the basis of marker gene expression. Pseudocells were also coloured by species and diffusion map based on MGE neuron developmental pseudotimes. **m**, *t*-SNE plots coloured by marker gene expression and in situ hybridization images from the Allen Developing Mouse Brain Atlas (available from https://developingmouse.brain-map.org/) showing expression of *Dlx5*, *Isl1* and *Sox6* in the mouse developing ventral forebrain embryonic day 13.5 (E13.5). **n**, Human–chimpanzee ventral differentially expressed genes are largely shared along the dorsal forebrain developmental trajectories. **o**, Human–chimpanzee DE directionalities and magnitudes and DE gene detection rates on the two trajectories. DE directionalities and magnitudes are consistent on the dorsal and MGE trajectories, with most of the shared DE genes showing the highest human–chimpanzee expression divergence at NPC. DE genes specifically detected on one trajectory have the tendency of higher detection rates on the trajectory where human–chimpanzee differential expression is detected.

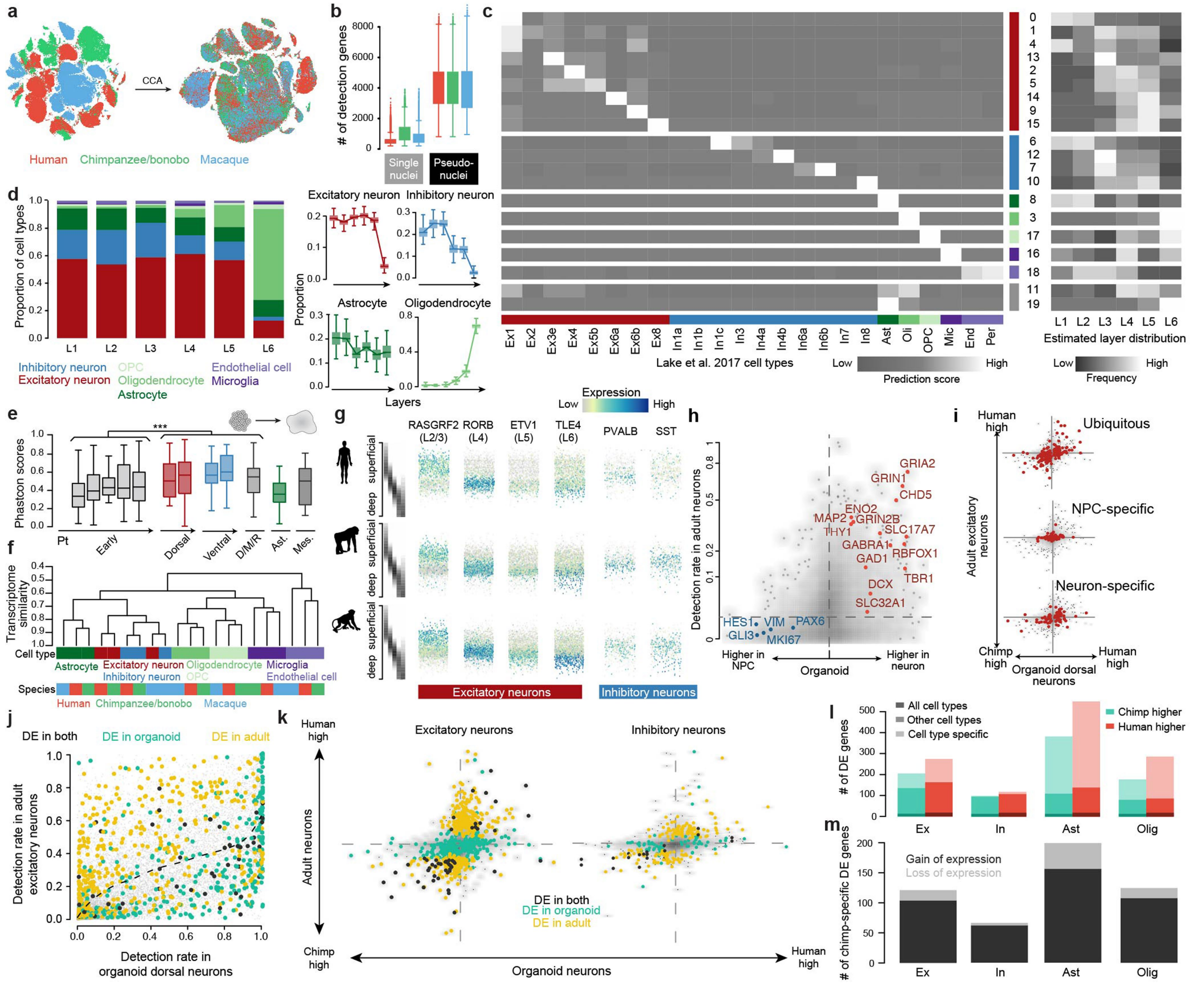


**Extended Data Fig. 8 | See next page for caption.**

**Extended Data Fig. 8 | Chromatin accessibility in cerebral organoids during development. a**, *t*-SNE projection of highly variable gene expression in Fluidigm C1-based scRNA-seq data of cerebral organoids. Cortical cells are coloured red, with larger points corresponding to cells with paired expression and chromatin accessibility data (data generated from the same cell suspension). A total of 94.4% (219 out of 232 cells) of cells with paired data are cortical, validating the cortical origins of the dissected cerebral organoid regions. **b**, Cerebral organoid accessible peaks are significantly and highly enriched (Fisher's exact test) for overlapping human VISTA enhancers active in the forebrain relative to all other tissues (left). Three representative human VISTA enhancers with validated activity in E11.5 mouse forebrain that overlap cerebral organoid peaks (out of 268 such enhancers) (right). **c**, Percentage of genes with accessible chromatin at the promoter of genes that are expressed or not expressed in human cerebral organoids. **d**–**f**, *t*-SNE projection of bias-corrected deviations in accessibility for 7-mers within organoid scATAC-seq peaks per cell, with cells coloured by cell state (NPC, neuron) for human (**d**, 221 cells), chimpanzee (**e**, 543 cells) and macaque (**f**, 118 cells). Binding motif deviation *Z*-scores for representative transcription factors are shown, as well as deviation *Z*-scores for overlapping DA snATAC-seq peaks in mouse developing forebrain excitatory neurons[51]. **g**, Signal intensity tracks of aggregated and individual single-cell chromatin accessibility data per cell state in human organoids at a NPC-specific promoter peak (left) and a neuron-specific promoter peak (right). For comparison, cerebral organoid bulk ATAC-seq chromatin accessibility data and human fetal brain bulk DNase-seq are shown. **h**, Enrichment of representative enriched biological process GO terms associated with human NPC DA peaks (gold) or human neuron DA peaks (light red) relative to all human organoid accessible peaks. Each point in the scatter plot represents a GO term and is coloured by their enrichment in NPCs (yellow), neurons (red), both (dark red) or neither (grey). **i**, *t*-SNE plots coloured by pseudotime, and heat maps showing binding motif deviation *Z*-scores for chosen transcription factors (rows) in all cells (columns) ordered in pseudotime for human (left) and chimpanzee (right). **j**, *t*-SNE projection of bias-corrected deviations in accessibility for 7-mers within scATAC-seq peaks per cell (518 cells), with cells coloured by time point, and organoid data coloured by cell state. Binding motif deviation *Z*-scores for representative transcription factors are shown to the right. **k**, *t*-SNE plot with cells coloured by their deviation *Z*-score for overlapping differentially accessible snATAC-seq peaks from mouse developing forebrain[51] radial glia cells (left) or excitatory neurons (right). **l**, Diffusion map projection using the top-250 differentially accessible peaks per time point or cell state. The principle curve fit through the cells is shown as a black line. **m**, Proportion of cells scaled by row for each time point or cell state over pseudotime. **n**, Heat map representing the deviation *Z*-score of transcription factor motifs that significantly vary over the time course plotted for each cell across pseudotime.



**Extended Data Fig. 9 | Chromatin accessibility differences in human and chimpanzee cerebral organoids. a**, Signal intensity tracks of aggregated single-cell and bulk chromatin accessibility data from human, chimpanzee and macaque at a human-specific NPC-specific DA peak (left) and a human-specific neuron-specific DA peak (right). **b**, The eight most significant human–chimp organoid DA peaks containing a fixed SNC and accessible only in the cerebral organoid stage that overlap a VISTA human enhancer with validated activity in the developing mouse forebrain (out of 68 such cases). For each DA peak, the accessibility across pseudotime is shown for human and chimpanzee with heat maps depicting cells where the peak is accessible (yellow) or inaccessible (black). The activity pattern of the overlapping VISTA enhancer in E11.5 mouse embryos is shown to the right. **c**, The proportion of DE genes (dark colour) or all expressed genes as background (light colour) with a human–chimp organoid DA peak overlapping the promoter region (blue) or distal to the promoter region (pink). The plot shows that DE genes between human and chimpanzee organoids are more likely to have a nearby DA peak than background. **d**, Fixed SNCs predicted to significantly alter transcription factor binding within human–chimp organoid DA peaks, with the name of the altered motif shown for peaks linked to DE genes (red points). On the right, signal intensity tracks for a human motif gain (top) and human motif loss (bottom) within a human–chimp DA peak. **e**, Altered transcription factor motifs grouped by family plotted for their alteration rate, which is the number of times a family member's motif is altered in human–chimp organoid DA peaks divided by the number of times it is detected in all accessible organoid peaks. **f**, Twenty transcription factors with the highest alteration rate, which is the number of times a motif is altered in human–chimp organoid DA peaks divided by the number of times it is detected in all accessible organoids peaks. Heat maps show their expression level in human and chimpanzee NPCs and neurons, with the bars to the left representing the average expression level across NPCs and neurons. **g**, Example of an accessible peak in chimpanzee and macaque that overlaps a computationally verified, non-polymorphic human conserved deletion (hCONDEL). **h**, **i**, Signal intensity tracks of aggregated single-cell or bulk chromatin accessibility data from human, chimpanzee and macaque for two genes, *LYPD1* (**h**) and *RAC1* (**i**), that have higher expression and specifically in humans, with genomic regions with gain of accessibility detected specifically in humans. Gene expression is shown in heat maps (bottom).



**Extended Data Fig. 10 | Supplementary analysis of human, chimpanzee and macaque adult brain snRNA-seq.** **a**, The snRNA-seq data of adult brains in human (50,035), chimpanzee and bonobo (33,847) and macaque (50,403) were integrated using Seurat v.3. **b**, Box plots (boxes show IQR and whiskers show 1.5 × IQR) showing the number of detected genes in single nuclei and pseudonuclei (3,420 human, 3,831 chimpanzee and 4,623 macaque pseudonuclei). **c**, Heat map showing the average prediction scores of each of the 20 identified clusters to each of the cell types reported[66] by Lake et al. (2016), as well as their estimated distributions in different cortical layers in humans. Clusters are grouped in major cell classes. **d**, Cell-type composition of layers and layer distribution of cell types in human. Left, stacked bars showing the estimated cell-type composition of different layers. Right, box plots (boxes show IQR and whiskers show 1.5 × IQR) showing the estimated proportion per layer for four cell classes: excitatory neurons, inhibitory neurons, astrocytes and oligodendrocytes. **e**, Genomic conservation based on average phastCon scores of developmental stage markers (in total 2,000 genes) from iPSCs to neurons in human cerebral organoids (***$P < 10^{-10}$, two-sided Wilcoxon's rank-sum test, $n_1 = 818$ genes, $n_2 = 188$ genes). **f**, Hierarchical clustering of the average transcriptome of seven cell classes in the three species. **g**, Expression of layer markers (*RASGRF2*, *RORB*, *ETV1* and *TLE4*) in excitatory neurons and inhibitory neuron subtype markers (*PVALB* and *SST*) in inhibitory neurons, along the predicted laminar origin of the pseudonuclei in human, chimpanzee and bonobo, and macaque. **h**, Detection rate in adult tissue of genes being differentially expressed between NPCs and neurons in organoids. **i**, Comparison of human–chimpanzee DE in adult excitatory neurons and that in organoid dorsal neurons for the robust DE genes detected in the organoid dorsal forebrain trajectory. Three categories of DE genes are highlighted: ubiquitous DE in organoids (top), DE only in NPCs (middle) and DE only in neurons (bottom). **j**, Comparison of gene-detection rates in organoid dorsal neurons and adult excitatory neurons, with human–chimpanzee DE genes in adult excitatory neurons coloured in yellow, DE genes in organoid dorsal neurons coloured in green, and shared DE genes coloured in black. The dashed curve shows the fitted relationship between the two systems using all genes. Area below the curve represents higher detection rate in organoid neurons than adult neurons and area above the curve represents higher detection rate in adult neurons. **k**, Comparison of human–chimpanzee DE (left) between organoid dorsal neurons and adult excitatory neurons, and between organoid ventral MGE neurons and adult inhibitory neurons (right). Densities are shown as grey scale shadows, with human–chimpanzee DE genes highlighted (yellow, DE only in adult; green, DE only in organoids; black, DE in both). **l**, Number of human and chimp DE genes for cell classes based on all cell types, a subset of cell types and specific cell types. **m**, Number of chimpanzee-specific DE genes across cell classes. The majority of the chimpanzee-specific DE genes have gain of expression (dark) rather than loss of expression (light).

nature research

Corresponding author(s): Barbara Treutlein, J. Gray Camp, Zhisong He

Last updated by author(s): Aug 29, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | See Methods. All codes are available upon request. |
| Data analysis | See Methods. All codes are available upon request. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data are deposited to EMBL-EBI ArrayExpress. The single-cell RNA-seq data based on 10X Genomics is deposited with the accession number E-MTAB-7552. The single-cell RNA-seq data based on Fluidigm C1/Smart-Seq2 is deposited with the accession number E-MTAB-8234. The single-cell ATAC-seq data is deposited with the accession numbers E-MTAB-8089 (human), E-MTAB-8043 (chimp), E-MTAB-8083 (bonobo), and E-MTAB-8087 (macaque). The bulk ATAC-seq data is deposited with the accession number E-MTAB-8228. The single-nucleus RNA-seq data of the adult brain samples is deposited with the accession number E-MTAB-8230. The bulk RNA-seq data is deposited with the accession number E-MTAB-8231. Data are also available for exploration via the public interactive browser scApeX (https://bioinf.eva.mpg.de/shiny/sample-apps/scApeX/)

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Samples are listed in Supplementary Table 1.<br>For determining the number of lines in DE analysis between human, chimp and macaque, see Extended Figure 8. For determining the number of cells to be sequenced per organoid and species, our project is the most extensive analysis thus far; based on Camp et al. (PNAS, 2015), Quadrato et al. (Nature, 2017) and Pollen et al. (Cell, 2019), we believe that we sufficiently sample the heterogeneity with the number of cells sequenced. |
| Data exclusions | We excluded low quality cells using criteria as described in the Methods. Besides that, no data were excluded. |
| Replication | We analyzed multiple organoids from multiple cell lines to determine the reproducibility of gene expression patterns across organoids and cell lines. We analyzed multiple organoids for different time points. We analyzed multiple individuals per species for the adult brain data. |
| Randomization | We try to mitigate any batch effects by analyzing organoids from multiple batches for a given line. In a set of experiments, we also multiplexed different cell lines from different species or different human individuals. We pooled tissue material from multiple individuals from different species for the adult brain data to reduce batch effects. |
| Blinding | Investigators were not blinded during data acquisition. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | See Methods, in the section "Immunohistochemistry". |
| Validation | See Methods, validations as provided by the manufacturer. |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | See Methods, in the section "Pluripotent stem cell lines and organoid culture". |
| Authentication | See Methods, in the section "Pluripotent stem cell lines and organoid culture". Cells were further authenticated based on single cell RNA-Seq reads compared to single nucleotide polymorphisms or species differences. |
| Mycoplasma contamination | Cell lines were tested for mycoplasma contamination on a regular basis using a PCR-based test and were found to be negative for mycoplasma. |
| Commonly misidentified lines (See ICLAC register) | None. |

## Flow Cytometry

### Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

| | |
|---|---|
| Sample preparation | See Methods, in the section "Single-nucleus and bulk RNA-Seq data generation" |
| Instrument | BD FACS Aria III and BD FACS Fusion. |
| Software | FACS Diva v.6.1.3 |
| Cell population abundance | Sorted nuclei populations were resorted using the same gating strategy and purity was above 90%. |
| Gating strategy | See methods, in the section "Single-nucleus and bulk RNA-Seq data generation". To distinguish DAPI positive and negative events, an unstained nuclei sample was used as a negative control to check the background signal and set the gate for the DAPI channel. |

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

https://doi.org/10.1038/s41586-019-1637-x

# A sensor kinase controls turgor-driven plant infection by the rice blast fungus

Lauren S. Ryder[1,2], Yasin F. Dagdas[1,2,3], Michael J. Kershaw[1], Chandrasekhar Venkataraman[4], Anotida Madzvamuse[4], Xia Yan[1,2], Neftaly Cruz-Mireles[2], Darren M. Soanes[1], Miriam Oses-Ruiz[1,2], Vanessa Styles[4], Jan Sklenar[2], Frank L. H. Menke[2] & Nicholas J. Talbot[1,2]*

**The blast fungus *Magnaporthe oryzae* gains entry to its host plant by means of a specialized pressure-generating infection cell called an appressorium, which physically ruptures the leaf cuticle[1,2]. Turgor is applied as an enormous invasive force by septin-mediated reorganization of the cytoskeleton and actin-dependent protrusion of a rigid penetration hypha[3]. However, the molecular mechanisms that regulate the generation of turgor pressure during appressorium-mediated infection of plants remain poorly understood. Here we show that a turgor-sensing histidine–aspartate kinase, Sln1, enables the appressorium to sense when a critical turgor threshold has been reached and thereby facilitates host penetration. We found that the Sln1 sensor localizes to the appressorium pore in a pressure-dependent manner, which is consistent with the predictions of a mathematical model for plant infection. A Δ*sln1* mutant generates excess intracellular appressorium turgor, produces hyper-melanized non-functional appressoria and does not organize the septins and polarity determinants that are required for leaf infection. Sln1 acts in parallel with the protein kinase C cell-integrity pathway as a regulator of cAMP-dependent signalling by protein kinase A. Pkc1 phosphorylates the NADPH oxidase regulator NoxR and, collectively, these signalling pathways modulate appressorium turgor and trigger the generation of invasive force to cause blast disease.**

Plant pathogenic fungi cause many of the world's most devastating crop diseases, and pose a constant threat to global food security[4,5]. The fungus *Magnaporthe oryzae* causes rice blast—the most widespread and serious disease of rice[1]—as well as wheat blast, which recently spread from South America to Bangladesh, threatening wheat production across South Asia[6,7].

To infect plants, *M. oryzae* develops a specialized infection cell called an appressorium[2] that ruptures the leaf cuticle using huge invasive force. The appressorium generates turgor of up to 8.0 MPa by accumulating high concentrations of glycerol and other polyols[8]. A differentiated cell wall that is rich in melanin is essential for the generation of turgor, acting as a rigid structural barrier to prevent the efflux of solutes[8,9]. The translation of appressorium turgor into mechanical force causes a narrow penetration hypha to emerge from the base of the appressorium and breach the cuticle of the rice leaf[1]. Septin GTPases form a toroidal, hetero-oligomeric complex at the appressorium pore, and this complex scaffolds cortical F-actin at the point of plant infection. Septins provide cortical rigidity and act as a diffusion barrier for polarity determinants that mediate membrane curvature and protrusion of the penetration hypha[3]. Septin-mediated reorientation of F-actin also requires the regulated synthesis of reactive oxygen species by NADPH oxidases (NOX)[10].

We set out to investigate how the internal pressure of the appressorium is modulated to control repolarization. We reasoned that the appressorium must reach a critical turgor threshold to trigger septin-mediated reorganization of the cytoskeleton. To test this idea, we first artificially lowered the turgor of the appressorium, and quantified the frequency of assembly of septin rings and the resulting disease lesions (Fig. 1). We observed fewer lesions when high concentrations of glycerol were applied to rice seedlings, demonstrating a relationship between appressorium turgor and infection (Fig. 1a, b). By contrast, application of glycerol to intact rice leaves had no effect (Extended Data Fig. 1). Septin organization was also impaired after treatment of appressoria with glycerol, and by treatment with the melanin biosynthesis inhibitor tricyclazole when applied before 16 hours post-inoculation (h.p.i.) (Fig. 1c, Extended Data Fig. 2a). Furthermore, septins and F-actin were mislocalized in the melanin-deficient mutants Δ*alb1*, Δ*rsy1* and Δ*buf1*, which fail to generate sufficient turgor for plant infection[9,11] (Fig. 1d, Extended Data Fig. 2b). We conclude that septin organization at the appressorium pore requires a critical threshold of cellular turgor and that this is essential for plant infection.

We postulated that a turgor sensor in appressoria must be necessary for the modulation of turgor pressure. To test this idea, we developed a mathematical model that couples geometric evolution laws for motion of the fungus and leaf surface with equations for the biosynthesis of melanin at the appressorium cortex, recruitment of septins and reorganization of F-actin (see Supplementary Information for a description and critical analysis of both the utility and the limitations of the mathematical model). A simulation of the model shows dynamics of appressorium repolarization that are consistent with experimental observations, and predicts that a mutant that lacks the turgor sensor will develop non-functional, hyper-melanized appressoria with excess turgor and aberrant deposition of septin and actin (Extended Data Fig. 3a, b; Supplementary Videos 1 and 2).

We next set out to identify the potential turgor sensor in *M. oryzae*. We noted that among potential candidates, *M. oryzae* possesses a homologue of the Sln1 histidine–aspartate sensor kinase—a known yeast osmosensor that modulates hyperosmotic adaptation through the high osmolarity glycerol (HOG) MAPK pathway[12]. The *HOG1* homologue in *M. oryzae*, *OSM1*, is dispensable for pathogenicity and glycerol production, suggesting that regulation of turgor is *OSM1*-independent[13]. *SLN1* was previously shown to be necessary for virulence in *M. oryzae*, but its function is unknown[14] (Fig. 2a, b, Extended Data Fig. 3c). Appressoria of a Δ*sln1* mutant generate extremely high turgor, as measured by incipient cytorrhysis[14] (Fig. 2c), but are non-functional. Live-cell imaging of *M. oryzae* that express a functional Sln1–GFP fusion showed that Sln1 localizes to the appressorium pore as infection cells generate pressure (Extended Data Fig. 3d, Supplementary Video 3). Localization of Sln1 was also sensitive to changes in turgor, and exposure to hyperosmotic glycerol led to Sln1 mislocalization (Fig. 2d) whereas the nuclear marker histone H1–GFP was unaffected (Extended Data Fig. 3e). The Δ*sln1* mutant also formed hyper-melanized appressoria (Fig. 2e)—a phenotype that was partially reversed by exposure to tricyclazole (Fig. 2e). Applying hyperosmotic stress to appressoria also enhanced the deposition of melanin in the

---

[1]Department of Biosciences, University of Exeter, Exeter, UK. [2]The Sainsbury Laboratory, University of East Anglia, Norwich, UK. [3]Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria. [4]School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Brighton, UK. *e-mail: nick.talbot@tsl.ac.uk

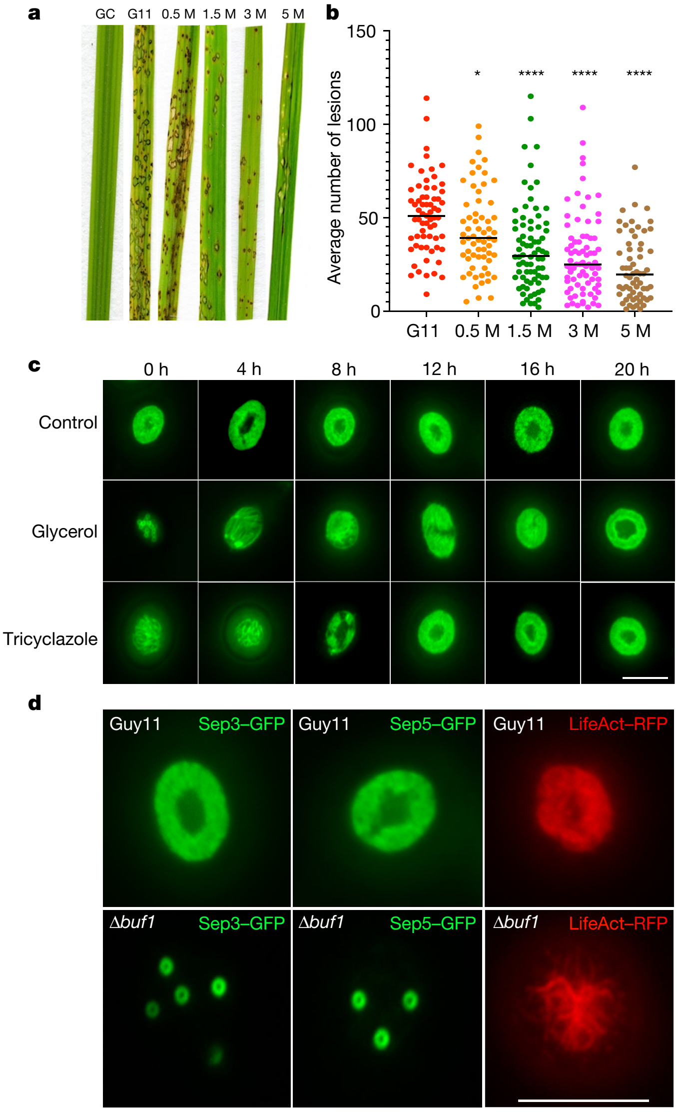


**Fig. 1 | Reducing appressorium turgor prevents septin-ring formation and impairs blast infection.** **a**, Seedlings of rice cultivar CO-39 were inoculated with a suspension of spores of *M. oryzae* ($1 \times 10^5$ conidia per ml in an aqueous solution of 0.2% gelatin) of the wild-type strain Guy11. At 5 h.p.i., seedlings were sprayed with glycerol solutions of 0.5 M, 1.5 M, 3 M and 5 M, respectively. GC, glycerol control; G11, Guy11 control. Seedlings were incubated for 5 d to observe the symptoms of blast disease. **b**, Dot plot showing the frequency of disease lesions observed in a 5-cm zone from each individual leaf harvested (60 leaves were harvested per treatment). Data are the mean and individual data points for $n = 3$ independent biological replicates. A two-tailed, unpaired Student's *t*-test with Welch correction was used for comparisons with the Guy11 control (*$P = 0.0312$, ****$P < 0.0001$). **c**, Cellular localization of Sep5–GFP at the appressorium pore of Guy11 after treatment with 1.5 M glycerol or 100 μM tricyclazole between 0 and 20 h.p.i., imaged at 24 h.p.i. Images are representative of $n = 3$ independent biological replicates. **d**, Organization of Sep3–GFP, Sep5–GFP and LifeAct–RFP in the appressorium pore of the melanin-deficient mutant Δ*buf1*. Images represent $n = 3$ independent biological replicates. Scale bars, 10 μm (**c**, **d**).

wild-type Guy11 strain of *M. oryzae* (Extended Data Fig. 4), suggesting that Sln1 modulates melanization of the appressorium once sufficient turgor has been generated.

To identify cellular functions that are controlled by Sln1, we used RNA sequencing (RNA-seq) to compare global changes in gene expression in a Δ*sln1* mutant to wild-type Guy11 during appressorium development. Among 1,982 genes that were affected in expression by loss of *SLN1* at 16 h.p.i., the melanin biosynthetic genes *RSY1* and *BUF1* were significantly upregulated in the Δ*sln1* mutant—consistent with increased melanization (Fig. 2e, Extended Data Fig. 5a). Furthermore, Sln1–GFP was mislocalized in Δ*alb1*, Δ*rsy1* and Δ*buf1* mutants (Fig. 2f). Notably, the *M. oryzae* response regulator mutants Δ*ssk1* and Δ*skn7* also display enhanced melanization[15] and in *Cryptococcus neoformans*, Δ*skn7*, Δ*ssk1* and Δ*tco1* mutants show similar phenotypes[16]. We reasoned that Sln1 negatively regulates melanin biosynthesis and turgor generation, and triggers repolarization of the appressorium. We therefore tested whether the organization of the septin ring and toroidal F-actin network in appressoria was affected in Δ*sln1* mutants Sep5–GFP and LifeAct–RFP (a marker of F-actin) were both mislocalized in Δ*sln1* mutants (Fig. 2g, Extended Data Fig. 5b). Septin organization was also impaired in Δ*alb1* mutants (Extended Data Fig. 5c), and disrupted in Δ*buf1* mutants, which are blocked at a later stage in production of 1, 8-dihydroxynaphthalene (DHN)-melanin (Fig. 1d). The Δ*sln1* mutant therefore continues to generate turgor in the appressorium, but reorganization of septins and formation of the penetration peg do not occur.

To investigate the putative Sln1 turgor-sensing complex, we immunoprecipitated Sln1–GFP from appressorium protein extracts at 16 h.p.i. and performed liquid chromatography–tandem mass spectrometry (LC–MS/MS) (Fig. 3a). Sln1 putatively interacts with two mechanosensitive ion-channel proteins, Mic1 and Mic3—similarly to previously described yeast proteins that respond to osmotic shock[17]. Blocking mechanosensitive ion channels with gadolinium and verapamil prevented appressorium formation (Extended Data Fig. 6), and although Mic1, Mic2 and Mic3 were individually dispensable for virulence, Mic2–GFP localized to the centre of the appressorium pore in an *SLN1*-dependent manner (Extended Data Fig. 7a–c). Sln1 also putatively interacts with two chitin synthases that are required for biosynthesis of the fungal cell wall, Chs4 and Chs5, and staining Δ*sln1* with calcofluor white revealed aberrant deposition of chitin within the cell wall of the appressorium (Extended Data Fig. 8a). This mirrors a previous study in *Arabidopsis thaliana*, which showed that *TOD1*—an alkaline ceramidase that regulates the turgor of guard cells and pollen tubes—acts by regulating cell-wall remodelling[18]. In addition, Sln1 interacts with Sum1, the regulatory subunit of cAMP-dependent protein kinase A (PKA), in both co-immunoprecipitation and yeast two-hybrid analyses (Fig. 3a, Extended Data Fig. 8b). PKA regulates the mobilization of lipid bodies and lipolysis, which leads to glycerol-dependent generation of turgor in *M. oryzae*[19]. The expression of *SUM1* is increased in a Δ*sln1* mutant, suggesting that Sln1 negatively regulates the PKA pathway to modulate the biosynthesis of glycerol (Fig. 3b). Consistent with this idea, the PKA drug inhibitor H-89 disrupts the organization of the appressorium pore in a dose-dependent manner, and localization of Sep5–GFP, gelsolin–GFP and Sln1–GFP is also impaired in a Δ*cpka* mutant (GenBank accession Q01143; Fig. 3c, Extended Data Fig. 9a–c). CpkA–GFP localizes to the appressorium pore (Fig. 3d) during the onset of turgor, consistent with its increased gene expression at this time (Fig. 3e).

Sln1 can also interact with protein kinase C (Pkc1), the central regulator of the cell-integrity pathway (Fig. 3b, Extended Data Fig. 8b). *PKC1* is an essential gene in *M. oryzae*, so to test its function in appressorium repolarization we used an allelic replacement mutant, *PKC1*$^{AS}$, which expresses an analogue-sensitive (Shokat) version of the kinase that is specifically sensitive to the ATP analogue 1NA-PP1[20]. Inhibition of Pkc1 by 1NA-PP1 disrupted the organization of Sep3–GFP, LifeAct–RFP and gelsolin–GFP at the appressorium pore (Fig. 3f), which was reversed by removal of 1Na-PP1 (Extended Data Fig. 10a). RNA-seq analysis of the *PKC1*$^{AS}$ mutant in the presence or absence of 1NA-PP1 also showed a significant reduction in the expression of *NOX1*, *NOX2* and *NOXR* after 24 h[20] (Extended Data Fig. 10b). Furthermore, yeast two-hybrid analysis revealed transient interactions between Nox1, Nox2, NoxR and Pkc1 (Extended Data Fig. 10c). This is consistent with studies in humans that have demonstrated that PKC is required for phosphorylation of gp91phox (Nox2), and that this phosphorylation enhances the diaphorase activity of gp91phox and its binding to

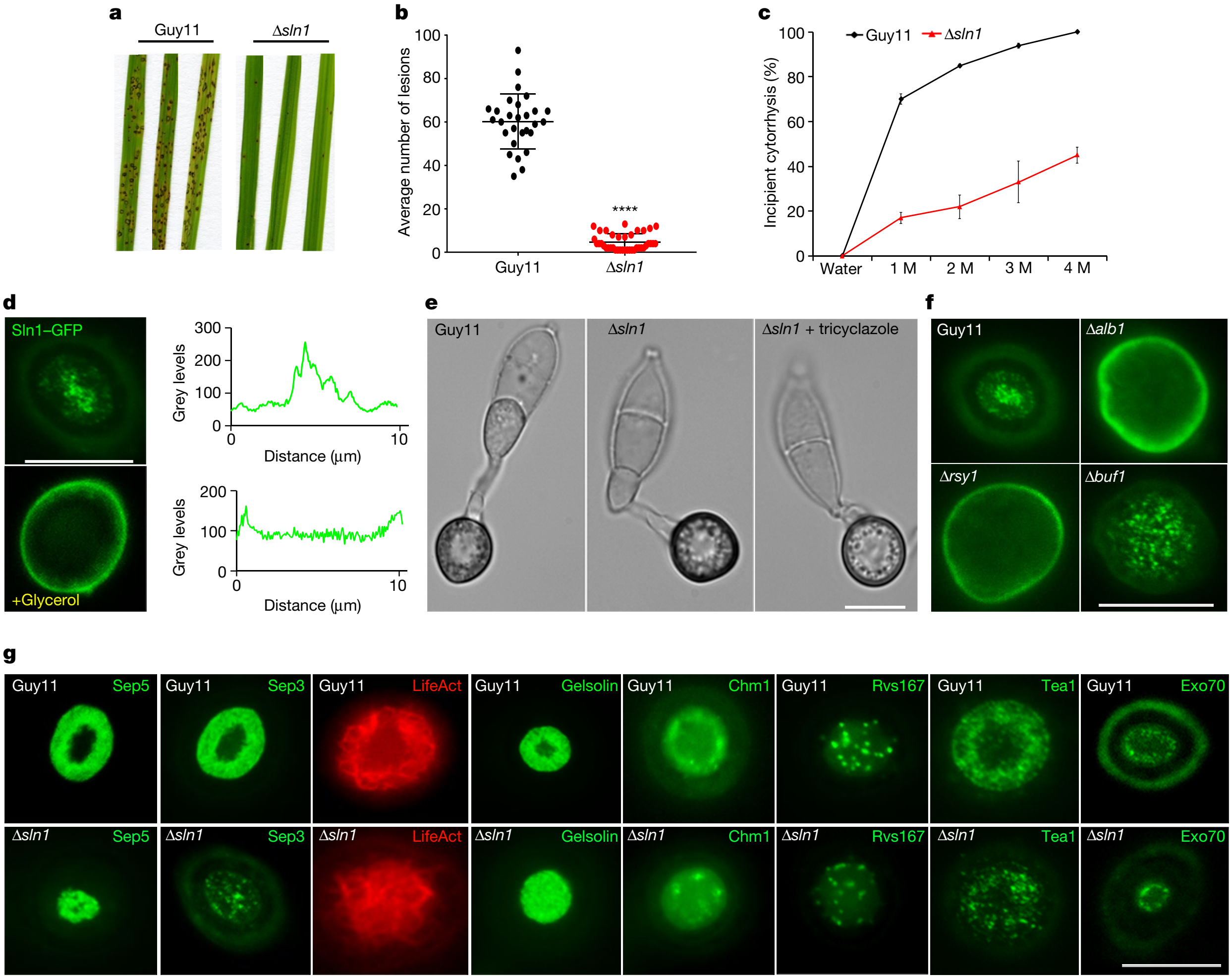


**Fig. 2 | Identification of the Sln1 turgor-sensing kinase in *M. oryzae*.** **a**, Rice blast assay of a Δ*sln1* kinase mutant. Rice cultivar CO-39 was inoculated with a 0.2% gelatin suspension of $1 \times 10^5$ conidia per ml of wild-type Guy11 or the isogenic Δ*sln1* mutant, and incubated for 120 h to allow development of the symptoms of blast disease. **b**, Dot plot showing the frequency of disease lesions observed in a 5-cm zone from each individual leaf harvested (28 individual leaves were harvested per strain). Data are mean ± s.e.m. and individual data points for $n = 2$ independent biological replicates. $^{****}P < 0.0001$ (two-tailed, unpaired Student's *t*-test with Welch correction). **c**, Percentage of Guy11 and Δ*sln1*-mutant appressoria that undergo incipient cytorrhysis after treatment with glycerol solutions of 1.0–4.0 M. Data are mean ± s.e.m. for $n = 3$ independent experiments; 50 appressoria were counted per experiment. $P < 0.01$ for Guy11 versus 0.5 M glycerol; $P < 0.0001$ for Guy11 versus 1.5 M, 3 M and 4 M glycerol (two-tailed unpaired Student's *t*-test). **d**, Left, epifluorescence micrographs showing the cellular distribution of Sln1–GFP in appressoria at 24 h.p.i., after exposure of appressoria to 1.5 M glycerol at 5 h.p.i. Right, line-scan graphs showing Sln1–GFP fluorescence in transverse sections of individual appressoria. Images are representative of $n = 3$ independent repeats of the experiment. **e**, Micrographs of appressoria of Guy11 and a Δ*sln1* mutant to show the melanin layer. Hyper-melanization of Δ*sln1* could be reversed by exposure to tricyclazole. Images are representative of $n = 3$ independent repeats of the experiment. **f**, Sln1–GFP expression and localization in appressoria of Δ*alb1*, Δ*rsy1* and Δ*buf1* mutants at 24 h.p.i. Images are representative of $n = 3$ independent repeats of the experiment. **g**, Localization patterns of Sep3-GFP, Sep5-GFP, LifeAct–RFP, gelsolin–GFP, Chm1–GFP, Tea1–GFP and Exo70–GFP in appressorium pores of Guy11 and a Δ*sln1* mutant. Images are representative of $n = 3$ independent repeats of the experiment. Scale bars, 10 μm (**d**–**g**).

Rac2, p67phox (NoxR) and p47phox (Bem1)[21]. By phosphoproteomic analysis, we observed that Pkc1 phosphorylates NoxR at serine 321 (Extended Data Fig. 10d, Supplementary Table 2), consistent with activation of the NADPH oxidase complex[10,22] (which is necessary for septin-dependent plant infection) by Pkc1. Notably, Pkc1 also phosphorylates the phosphodiesterase PdeH—which regulates the PKA pathway—at serine 883, in addition to phosphorylating other proteins that are predicted to be involved in the sensing of turgor (Supplementary Table 2). Incipient cytorrhysis of a Δ*pdeH* mutant shows that it generates excess appressorium turgor, suggesting that PdeH is regulated by Sln1 (Extended Data Fig. 10e). PdeH in *M. oryzae* was previously shown to mediate cross-talk between the PKA, HOG and cell-integrity pathways[23]—consistent with a role in turgor sensing.

Finally, a cell-cycle checkpoint, triggered by the generation of appressorium turgor and by melanization, is known to regulate septin-dependent infection[24]. We therefore blocked the progression of cells into S phase by treatment with hydroxyurea, which prevented Sln1 recruitment to the appressorium pore (Extended Data Fig. 10f). Operation of this cell-cycle checkpoint is thus critical to the sensing of turgor in appressoria.

When considered together with our mathematical modelling, the experimental data presented here provide evidence that turgor-driven infection of plants by *M. oryzae* is controlled by a sensor kinase, Sln1 (Fig. 4). Once a threshold of turgor is reached, Sln1 negatively regulates the biosynthesis of melanin and production of glycerol as the appressorium nucleus enters S-phase[24]. Isotropic expansion of the pressurized appressorium ceases, and Sln1 acts through the Pkc1-dependent cell-integrity pathway to activate the Nox2–NoxR NADPH oxidase—thereby recruiting septins to the appressorium pore and reorganizing F-actin to facilitate force generation and polarized growth. Sln1 also inhibits the cAMP/PKA pathway; Pkc1 acts directly on PdeH to modulate levels of cAMP, and may induce

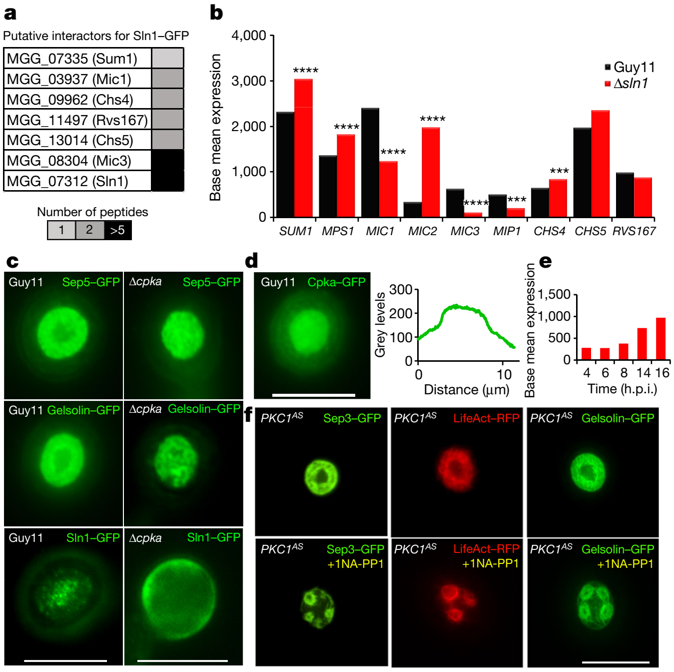


**Fig. 3 | Characterization of the Sln1 turgor-sensing complex in *M. oryzae*.** **a**, Putative Sln1-interacting peptides were immunoprecipitated from appressorium protein extracts at 16 h.p.i. from *M. oryzae* expressing Sln1–GFP or free cytoplasmic GFP using anti-GFP antibodies, and LC–MS/MS was performed to identify unique putatively interacting peptides. Colours represent the number of identified peptides for each selected protein. **b**, Differential expression of *SUM1*, *MPS1*, *MIC1*, *MIC2*, *MIC3*, *MIP1*, *CHS4*, *CHS5* and *RVS167* in Δ*sln1*-mutant appressoria compared to Guy11 appressoria at 16 h.p.i. by RNA-seq analysis. $n = 3$ independent biological repeats of the experiment for each strain. $***P < 0.001$, $****P < 0.0001$ (from multiple testing using the Benjamini–Hochberg method to estimate false discovery rate). **c**, Cellular localization of Sep5–GFP, gelsolin–GFP and Sln1–GFP in appressorium pores of Guy11 and a Δ*cpka* mutant at 24 h.p.i. Images are representative of $n = 3$ independent repeats of the experiment. **d**, Left, cellular distribution of Cpka–GFP in appressorium pores at 24 h.p.i. Right, line-scan graph showing Cpka–GFP fluorescence in a transverse section of an individual appressorium. Images are representative of $n = 3$ independent repeats of the experiment. **e**, Relative expression of *CPKA* from 4–16 h.p.i. during appressorium development. Data are from SuperSAGE analysis[26]. **f**, Localization of Sep3–GFP, LifeAct–RFP and gelsolin–GFP in appressorium pores of $PKC1^{AS}$ mutants in the presence or absence of 1NA-PP1. Images are representative of $n = 3$ independent repeats of the experiment.

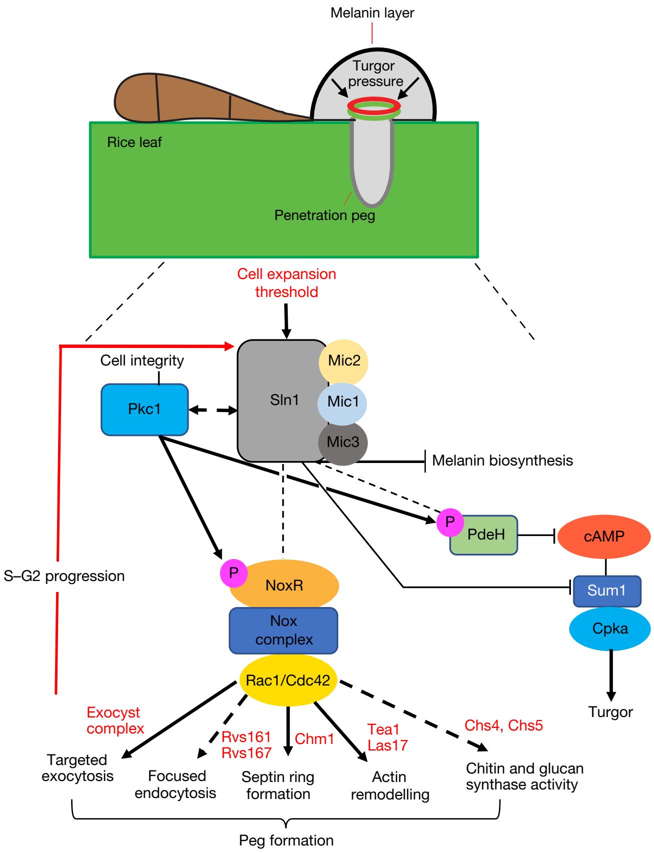


**Fig. 4 | Model of turgor-driven invasion of a plant cell by the rice blast fungus.** The Sln1 sensor kinase responds to appressorium turgor by interaction with a set of upstream monitors of cell expansion, including the stretch-activated ion-channel proteins Mic1, Mic2 and Mic3. Once a threshold of turgor is reached, Sln1 negatively regulates melanin biosynthesis and the cAMP/PKA pathway. Pkc1 acts directly on the PdeH phosphodiesterase to modulate levels of cAMP, and also acts to control lipolysis and glycerol production. Sln1 is then necessary for recruitment of septins to the appressorium pore, which requires Pkc1 and the NADPH oxidase Nox2. Septins tether cortical F-actin to the membrane, facilitating the formation of a toroidal network of F-actin, and organizing the exocyst complex and a large family of endocytic proteins at the pore. The septin ring acts as a diffusion barrier to ensure repolarization of the penetration peg, which involves the focused polymerization of F-actin and the activity of chitin and glucan synthases. A pressure-dependent S-phase checkpoint is also triggered[22] and is necessary for the action of Sln1. Collectively, these processes lead to breaching of the rice leaf cuticle.

glycerol efflux through the channel protein Mip1 (Fig. 3b). The septin ring acts as a diffusion barrier to ensure the localization of polarity determinants and regulate the polymerization of F-actin[3], recruitment of the exocyst complex[25] and activity of associated chitin and glucan synthases. Collectively, these processes lead to protrusion of a rigid penetration hypha, rupture of the rice leaf cuticle and onset of rice blast disease.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1637-x.

Received: 18 October 2017; Accepted: 4 September 2019;
Published online 9 October 2019.

1. Talbot, N. J. On the trail of a cereal killer: exploring the biology of *Magnaporthe grisea*. *Annu. Rev. Microbiol.* **57**, 177–202 (2003).
2. Ryder, L. S. & Talbot, N. J. Regulation of appressorium development in pathogenic fungi. *Curr. Opin. Plant Biol.* **26**, 8–13 (2015).
3. Dagdas, Y. F. et al. Septin-mediated plant cell invasion by the rice blast fungus, *Magnaporthe oryzae*. *Science* **336**, 1590–1595 (2012).
4. Pennisi, E. Sowing the seeds for the ideal crop. *Science* **327**, 802–803 (2010).
5. Fisher, M. C. et al. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**, 186–194 (2012).
6. Islam, M. T. et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biol*. **14**, 84 (2016).
7. Inoue, Y. et al. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science* **357**, 80–83 (2017).
8. de Jong, J. C. MCCormack, B. J., Smirnoff, N. & Talbot, N. J. Glycerol generates turgor in rice blast. *Nature* **389**, 244–245 (1997).
9. Chumley, F. G. & Valent, B. Genetic-analysis of melanin-deficient, nonpathogenic mutants of *Magnaporthe grisea*. *Mol. Plant Microbe Interact.* **3**, 135–143 (1990).
10. Ryder, L. S. et al. NADPH oxidases regulate septin-mediated cytoskeletal remodeling during plant infection by the rice blast fungus. *Proc. Natl Acad. Sci. USA* **110**, 3179–3184 (2013).
11. Wilson, R. A. & Talbot, N. J. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nat. Rev. Microbiol.* **7**, 185–195 (2009).
12. Tao, W., Deschenes, R. J. & Fassler, J. S. Intracellular glycerol levels modulate the activity of Sln1p, a *Saccharomyces cerevisiae* two-component regulator. *J. Biol. Chem*. **274**, 360–367 (1999).
13. Dixon, K. P., Xu, J.-R., Smirnoff, N. & Talbot, N. J. Independent signaling pathways regulate cellular turgor during hyperosmotic stress and appressorium-mediated plant infection by *Magnaporthe grisea*. *Plant Cell* **11**, 2045–2058 (1999).

14. Zhang, H. et al. A two-component histidine kinase, *MoSLN1*, is required for cell wall integrity and pathogenicity of the rice blast fungus, *Magnaporthe oryzae*. *Curr. Genet*. **56**, 517–528 (2010).
15. Motoyama, T. et al. Involvement of putative response regulator genes of the rice blast fungus *Magnaporthe oryzae* in osmotic stress response, fungicide action, and pathogenicity. *Curr. Genet.* **54**, 185–195 (2008).
16. Bahn, Y. S., Kojima, K., Cox, G. M. & Heitman, J. A unique fungal two-component system regulates stress responses, drug sensitivity, sexual development, and virulence of *Cryptococcus neoformans*. *Mol. Biol. Cell* **17**, 3122–3135 (2006).
17. Nakayama, Y., Hirata, A. & Iida, H. Mechanosensitive channels Msy1 and Msy2 are required for maintaining organelle integrity upon hypoosmotic shock in *Schizosaccharomyces pombe*. *FEMS Yeast Res*. **14**, 992–994 (2014).
18. Chen, L. Y. et al. The *Arabidopsis* alkaline ceramidase TOD1 is a key turgor pressure regulator in plant cells. *Nat. Commun*. **6**, 6030 (2015).
19. Thines, E., Weber, R. W. S. & Talbot, N. J. MAP kinase and protein kinase A-dependent mobilization of triacylglycerol and glycogen during appressorium turgor generation by *Magnaporthe grisea*. *Plant Cell* **12**, 1703–1718 (2000).
20. Penn, T. J. et al. Protein kinase C is essential for viability of the rice blast fungus *Magnaporthe oryzae*. *Mol. Microbiol*. **98**, 403–419 (2015).
21. Raad, H. et al. Regulation of the phagocyte NADPH oxidase activity: phosphorylation of gp91$^{phox}$/NOX2 by protein kinase C enhances its diaphorase activity and binding to Rac2, p67$^{phox}$, and p47$^{phox}$. *FASEB J*. **23**, 1011–1022 (2009).
22. Mithoe, S. C. et al. Attenuation of pattern recognition receptor signaling is mediated by a MAP kinase kinase kinase. *EMBO Rep*. **17**, 441–454 (2016).
23. Yin, Z. et al. Phosphodiesterase MoPdeH targets MoMck1 of the conserved mitogen-activated protein (MAP) kinase signalling pathway to regulate cell wall integrity in rice blast fungus *Magnaporthe oryzae*. *Mol. Plant Pathol*. **17**, 654–668 (2016).
24. Osés-Ruiz, M., Sakulkoo, W., Littlejohn, G. R., Martin-Urdiroz, M. & Talbot, N. J. Two independent S-phase checkpoints regulate appressorium-mediated plant infection by the rice blast fungus *Magnaporthe oryzae*. *Proc. Natl Acad. Sci. USA* **114**, E237–E244 (2017).
25. Gupta, Y. K. et al. Septin-dependent assembly of the exocyst is essential for plant infection by *Magnaporthe oryzae*. *Plant Cell* **27**, 3277–3289 (2015).
26. Soanes, D. M., Chakrabarti, A., Paszkiewicz, K. H., Dawe, A. L. & Talbot, N. J. Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog*. **8**, e1002514 (2012).

## METHODS

**Fungal strains, growth conditions and DNA analysis.** Growth, maintenance and storage of *M. oryzae* isolates, medium composition, nucleic acid extraction and transformation were all as previously described[27]. Gel electrophoresis, restriction enzyme digestion, gel blots and sequencing were performed using standard procedures[28].

**Assays of appressorium development and plant infection, quantification of melanin thickness and live cell imaging of cytoskeletal components of *M. oryzae*.** Appressorium development was induced in vitro on borosilicate 18 × 18-mm glass coverslips (Thermo Fisher Scientific), adapted from a previous study[29]. A total of 50 μl of conidial suspension ($5 \times 10^4 \text{ ml}^{-1}$) was placed on a coverslip and incubated at 24 °C. Rice leaf sheaths were inoculated[30] to observe the development of invasive hyphae. The transgenic line expressing LTi6B was used to observe plant cell viability[31]. At the desired time points, tricyclazole (100 μM), an agent that inhibits melanin biosynthesis, was added to *M. oryzae* and infection-related development assayed. Glycerol was used at a final concentration of 1.5 M for coverslip assays unless otherwise stated, and 1–5 M for plant spraying, and samples were incubated at 24 °C. Gadolinium (100 μM) and verapamil (100 μM) were added at the indicated times to evaluate their effect on infection-related development (0–20 h). To determine the thickness of the melanin layer of the appressorium, appressoria were sampled at random intervals at the cell cortex. Development of appressoria was observed using an IX81 motorized inverted microscope (Olympus) and images were captured using a Photometrics CoolSNAP HQ2 camera (Roper Scientific), under control of the Metamorph software package (MDS Analytical Technologies). Datasets were compared using an unpaired Student's *t*-test.

**Generation of GFP fusion plasmids.** DNA sequences were retrieved from the *M. oryzae* database (http://fungi.ensembl.org/Magnaporthe_oryzae/Info/Index) and used to design primers (Supplementary Table 1). In-Fusion cloning based on in vitro homologous recombination was performed to generate Sln1–GFP and Cpka–GFP, using a commercial kit (In-Fusion Cloning kit; Clontech Laboratories). The primers used are shown in Supplementary Table 1. Sln1–GFP, Cpka–GFP and Mic2–GFP were inserted as EcoRI/HindIII fragments into a modified Strataclone (Stratagene) vector containing the *BAR* gene that confers bialophos (BASTA) resistance[32]. In all cases, several independent *M. oryzae* transformants were first screened for consistency of the fluorescent signal and a representative transformant was then selected for further analysis. Three independent experiments were performed in each case unless otherwise stated.

**Targeted gene deletion of *MIC1*, *MIC2* and *MIC3* in *M. oryzae*.** Targeted gene replacement of *M. oryzae MIC1*, *MIC2* and *MIC3* was performed using the split marker strategy[6]. To amplify the split *HPH*, *IVL1* or *BAR* templates, the primers used were M13 (forward) with HY/ VL/BA and M13 (reverse) with YG/IV/AR, as described[33]. *M. oryzae* mechanosensitive-ion-channel genes that contain the pfam domain (PF00924) were aligned with *Schizosaccharomyces pombe*, *Aspergillus nidulans* and *Neurospora crassa* (*Saccharomyces cerevisiae* and *Candida albicans* do not have genes with this domain). The sequence data for each mechanosensitive-ion-channel gene in *M. oryzae* were retrieved from the *M. oryzae* genome database at https://fungi.ensembl.org/Magnaporthe_oryzae/Info/Annotation/) and used to design specific primers (see Supplementary Table 1). Either Guy11 or the Δ*ku70* (*KU70* is also known as *YKU70*) mutant[33] was transformed with the deletion cassette for each gene fusion; *Mic1:hph*, *Mic2:ivl1* and *Mic3:bar* (2 μg of DNA for each flank). Transformants were selected in the presence of either hygromycin B (200 $\mu\text{g ml}^{-1}$), sulfonylurea (50 $\mu\text{g ml}^{-1}$) or bialophos (50 $\mu\text{g ml}^{-1}$) and were routinely screened and assessed using Southern blotting.

**Co-immunoprecipitation experiments and LC–MS/MS analysis.** Total protein was extracted from lyophilized *M. oryzae* appressoria (generated on borosilicate 18 × 18-mm glass coverslips (Thermo Fisher Scientific)) at 16 h.p.i., collected using a razor blade and snap-frozen in liquid nitrogen. *M. oryzae* strains that express Sln1–GFP and ToxA–GFP (control) were co-immunoprecipitated using the GFP-Trap protocol according to the manufacturer's instructions (ChromoTek). Preparation of peptides for LC–MS/MS was performed as follows. Proteins were separated by SDS–PAGE. Gels were cut into slices (5–10 mm) and proteins contained within gel slices were prepared for LC–MS/MS as described previously[34]. LC–MS/MS analysis was performed with an Orbitrap Fusion trihybrid mass spectrometer (Thermo Fisher Scientific) and a nanoflow high-performance liquid chromatography (HPLC) system (Dionex Ultimate3000, Thermo Fisher Scientific), as described previously[35] but with the following differences: MS/MS peak lists were exported in the Mascot generic file format using Discoverer v2.2 (Thermo Fisher Scientific). The database was searched with Mascot v.2.3 (Matrix Science), with the following differences: (1) the database searched with Mascot v.2.3 (Matrix Science) was the *M. oryzae* protein database with the inclusion of sequences of common contaminants such as keratins and trypsin; (2) carbamidomethylation of cysteine residues was specified as a fixed modification, and oxidized methionine was allowed as a variable modification. The other Mascot parameters used were as follows: (1) mass values were monoisotopic and the protein mass was unrestricted; (2) the peptide mass tolerance was 5 ppm and the fragment mass tolerance was 0.6 Da; (3) two missed cleavages were allowed with trypsin. All Mascot searches were collated and verified with Scaffold (Proteome Software), and the subset database was searched with the Mascot server v.2.4.1 (Matrix Science). Accepted proteins passed the following threshold in Scaffold: 95% confidence for protein match and minimum of two unique peptide matches with 95% confidence.

**Yeast two-hybrid analysis.** In-Fusion Cloning based on in vitro homologous recombination was performed to generate vectors that express Sln1, Nox1, Nox2 and NoxR in the pGADT7 prey vector, and Mps1, Pkc1[20] and Sum1 in the pGBKT7 bait vector. Genes were amplified from *M. oryzae* cDNA derived from mycelium grown on liquid Complete Medium (CM) using primers with a 15-bp overhang and a restriction site complementary to the target vector (Supplementary Table 1). Fragments were cloned into pGBKT7 and pGADT7 plasmids and linearized by digestion with BamHI and EcoRI. Yeast two-hybrid assays using pGADT7- or pGBKT7-based constructs (Clontech) were performed according to the manufacturer's instructions (MATCHMAKER Gold Yeast Two-Hybrid System).

**Comparative RNA-seq analysis.** Total RNA was extracted from appressoria of the wild-type strain Guy11 and Δ*sln1* null mutant at 16 h.p.i., which were developed on hydrophobic coverslips using the RNeasy Plant Mini Kit for Total RNA extraction (Qiagen). RNA-seq libraries were then prepared from 5 μg of total RNA with the Illumina TruSeq RNA Sample Preparation Kit (Agilent) according to the manufacturer's instructions. Libraries were sequenced using the Illumina HiSeq 2500 platform. Reads were aligned against version 8.0 of the *M. oryzae* genome using TopHat software and analysis of the data was performed using DESeq through moderated $\log_2$-transformed fold-change values (mod_lfc)[36,37]. Transcript abundances for each gene and adjusted *P* values and transcript abundance were determined as previously described[38].

**Protein extraction and phosphoproteomic enrichment.** Mycelium of the *M.oryzae* $PKC1^{AS}$ mutant and Guy11 was prepared from CM shake cultures (125 r.p.m.) at 24 °C for 48 h. Mycelium was filtered through miracloth (Calbiochem), divided and treated with 1NA-PP1 at a final concentration of 500 nM for 4 h. An untreated control was also prepared at the same time point. Mycelium was then filtered, washed in distilled water and frozen in liquid nitrogen.

Frozen tissue was ground to a fine powder in liquid nitrogen, resuspended in extraction buffer (8 M urea, 150 mM NaCl, 100 mM Tris pH 8, 5 mM EDTA, 1 $\mu\text{g ml}^{-1}$ aprotinin, 2 $\mu\text{g ml}^{-1}$ leupeptin) and mechanically disrupted (8 min, 1,000 r.p.m.) in a 30-ml Potter-Elvehjem homogenizer incubated on ice[22]. The homogenate was then fractionated by centrifugation for 30 min at 10,000*g* (Sorvall SW34 rotor). The supernatant was removed and then centrifuged for 60 min at 100,000*g* (Sorvall T-647.5 rotor) to separate the cytosolic (supernatant) and microsomal (pellet) fractions. The microsomal pellet was then resuspended in extraction buffer. For phosphopeptide enrichment, sample preparation started with 1–3 mg of cytosolic or microsomal protein extract (confirmed by Bradford assay) dissolved in bicarbonate buffer containing 8 M urea. First, protein extracts were reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP) for 30 min at 30 °C with gentle shaking, followed by alkylation of cysteine residues with 40 mM iodoacetamide at room temperature for 1 h. Samples were diluted to a final concentration of 1.6 M urea with 50 mM ammonium bicarbonate and digested overnight with trypsin (Promega; 1:100 enzyme to substrate ratio). Peptide digests were purified using C18 SepPak columns (Waters) as described[39]. Phosphopeptides were enriched using titanium dioxide ($TiO_2$; GL Science) with phthalic acid as a modifier[22]. Finally, phosphopeptides were eluted by a pH shift to 10.5 and immediately purified using C18 microspin columns (The Nest Group; loading capacity of 5–60 μg). After purification, all samples were desiccated in a speed-vac, stored at −80 °C and resuspended in 2% acetonitrile with 0.1% trifluoroacetic acid before mass-spectrometry analysis[22].

**Mass-spectrometry analysis of phosphopeptide-enriched samples.** LC–MS/MS analysis was performed using an Orbitrap Fusion trihybrid mass spectrometer (Thermo Fisher Scientific) and a nanoflow ultra-high-performance liquid chromatography (UHPLC) system (Dionex Ultimate 3000, Thermo Fisher Scientific). Peptides were trapped on a reverse-phase trap column (Acclaim PepMap, C18 5 μm, 100 μm × 2 cm, Thermo Fisher Scientific). Peptides were eluted in a gradient of 3–40% acetonitrile in 0.1% formic (solvent B) acid over 120 min, followed by a gradient of 40–80% B over 6 min at a flow rate of 200 $\text{nl min}^{-1}$ at 40 °C. The mass spectrometer was operated in positive-ion mode with a nano-electrospray ion source with ID 0.02-mm fused silica emitter (New Objective). A voltage of 2,200 V was applied via platinum wire held in PEEK T-shaped coupling union with the transfer capillary temperature set to 275 °C. The Orbitrap mass spectrometry scan resolution of 120,000 at 400 *m*/*z*, range 300–1,800 *m*/*z* was used, and the automatic gain control was set to $2 \times 10^5$ and the maximum injection time to 50 ms. In the linear ion trap, MS/MS spectra were triggered using a data-dependent acquisition method, with 'top speed' and 'most intense ion' settings. The selected precursor ions were fragmented sequentially both in the ion trap using collision-induced dissociation (CID) and in the higher-energy collisional dissociation (HCD) cell. Dynamic exclusion was set to 15 s. The charge state allowed between 2+ and 7+ charge states to be selected for MS/MS fragmentation.

Peak lists in the format of Mascot generic files (.mgf files) were prepared from raw data using the MSConvert package (Matrix Science). Peak lists were searched on Mascot server v.2.4.1 (Matrix Science) against either the *Magnaporthe oryzae* (isolate 70-15, v.8) database, or an in-house contaminants database. Tryptic peptides with up to two possible miscleavages and the charge states $+2$, $+3$, $+4$ were allowed in the search. The following modifications were included in the search: oxidized methionine; phosphorylation on serine, threonine, or tyrosine as a variable modification; and carbamidomethylated cysteine as a static modification. Data were searched with a monoisotopic precursor and fragment-ions mass tolerance set at 10 ppm and 0.6 Da, respectively. Mascot results were combined in Scaffold v.4 (Proteome Software) and exported in Excel (Microsoft Office)[22].

**Statistical analysis.** All experiments were conducted with technical and biological replicates at an appropriate sample size that was estimated on the basis of our previous experience. No statistical methods were used to predetermine sample size. No methods of randomization were applied but blinding was applied to the data on disease symptoms, which are shown in Fig. 1b. All experiments were replicated independently at least once, as indicated in each figure legend. Dot plots were routinely used to show individual data points for each experimental observation, and bar graphs, where shown, also contained individual data points for each experimental replicate. Statistical analyses were performed using GraphPad Prism 8 or Microsoft Excel. $P$ values $<0.05$ were considered significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. $P$ values $>0.05$ were considered non-significant and exact values are shown where appropriate. The sample sizes and statistical tests used are stated in each figure legend.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

All strains generated and datasets analysed during the current study, including codes and algorithms, are available either in public repositories as stated, or from the corresponding author on reasonable request.

27. Talbot, N. J., Salch, Y. P., Ma, M. & Hamer, J. E. Karyotypic variation within clonal lineages of the rice blast fungus, *Magnaporthe grisea*. *Appl. Environ. Microbiol.* **59**, 585–593 (1993).
28. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
29. Hamer, J. E. et al. A mechanism for surface attachment in spores of a plant pathogenic fungus. *Science* **239**, 288–290 (1988).
30. Kankanala, P., Czymmek, K. & Valent, B. Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. *Plant Cell* **19**, 706–724 (2007).
31. Mentlak, T. A. et al. Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. *Plant Cell* **24**, 322–335 (2012).
32. Lindsay, R. J., Kershaw, M. J., Pawlowska, B. J., Talbot, N. J. & Gudelj, I. Harbouring public good mutants within a pathogen population can increase both fitness and virulence. *eLife* **5**, e18678 (2016).
33. Kershaw, M. J. & Talbot, N. J. Genome-wide functional analysis reveals that infection-associated fungal autophagy is necessary for rice blast disease. *Proc. Natl Acad. Sci. USA* **106**, 15967–15972 (2009).
34. Petre, B., Win, J., Menke, F. L. H. & Kamoun, S. Protein–protein interaction assays with effector–GFP fusions in *Nicotiana benthamiana*. *Methods Mol. Biol.* **1659**, 85–98 (2017).
35. Bender, K. W. et al. Autophosphorylation-based calcium ($Ca^{2+}$) sensitivity priming and $Ca^{2+}$/calmodulin inhibition of *Arabidopsis thaliana* $Ca^{2+}$-dependent protein kinase 28 (CPK28)*J. Biol. Chem.* **292**, 3988–4002 (2017).
36. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
37. Trapnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
38. Soanes, D. M., Chakrabarti, A., Paszkiewicz, K. H., Dawe, A. L. & Talbot, N. J. Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PloS Pathog.* **8**, e1002514 (2012).
39. Ludwig, C., Claassen, M., Schmidt, A. & Aebersold, R. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol. Cell. Proteomics* **11**, M111.013987 (2012).
40. Chojnacki, S., Cowley, A., Lee, J., Foix, A. & Lopez, R. Programmatic access to bioinformatics tools from EMBL–EBI update: 2017. *Nucleic Acids Res.* **45**, W550–W553 (2017).

**Acknowledgements** We acknowledge technical support from O. Goode and T. Penn. This work was funded by a European Research Council (ERC) Advanced Investigator Award to N.J.T. under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC grant no. 294702 GENBLAST. L.S.R. acknowledges the late P. Ryder for support and encouragement.

**Author contributions** L.S.R., Y.F.D. and N.J.T. conceived and designed the project. L.S.R., Y.F.D., M.J.K., M.O.-R., N.C.-M. and X.Y. performed experimental work. D.M.S. performed bioinformatic analysis. C.V., A.M. and V.S. performed mathematical modelling. F.L.H.M., N.C.-M. and J.S. performed proteomic analysis. L.S.R. and N.J.T. wrote the paper with assistance and input of coauthors.

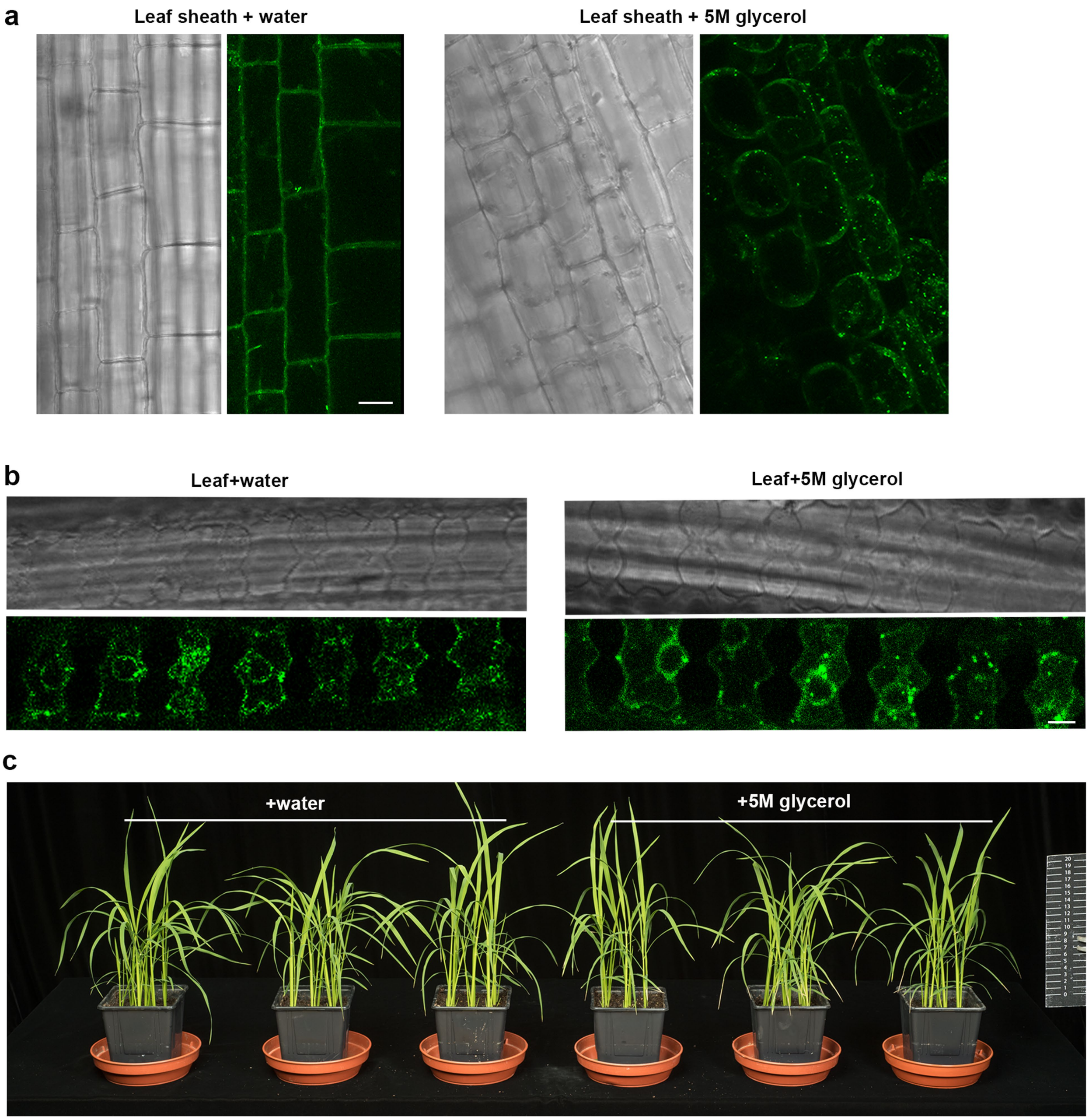
**Competing interests** The authors declare no competing interests.

**Additional information**

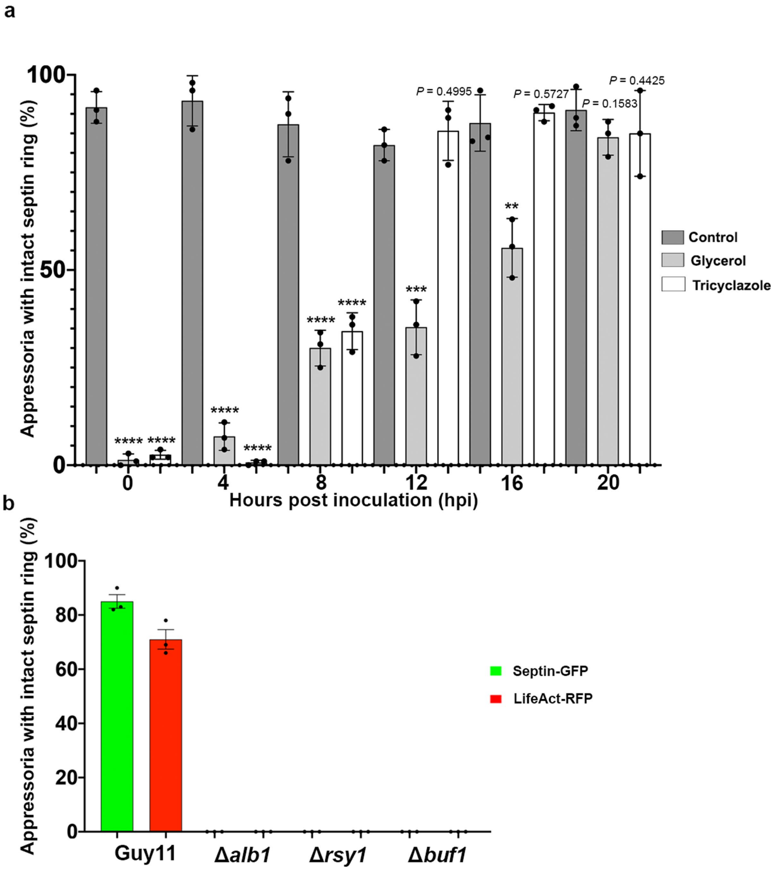
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1637-x.

**Correspondence and requests for materials** should be addressed to N.J.T.

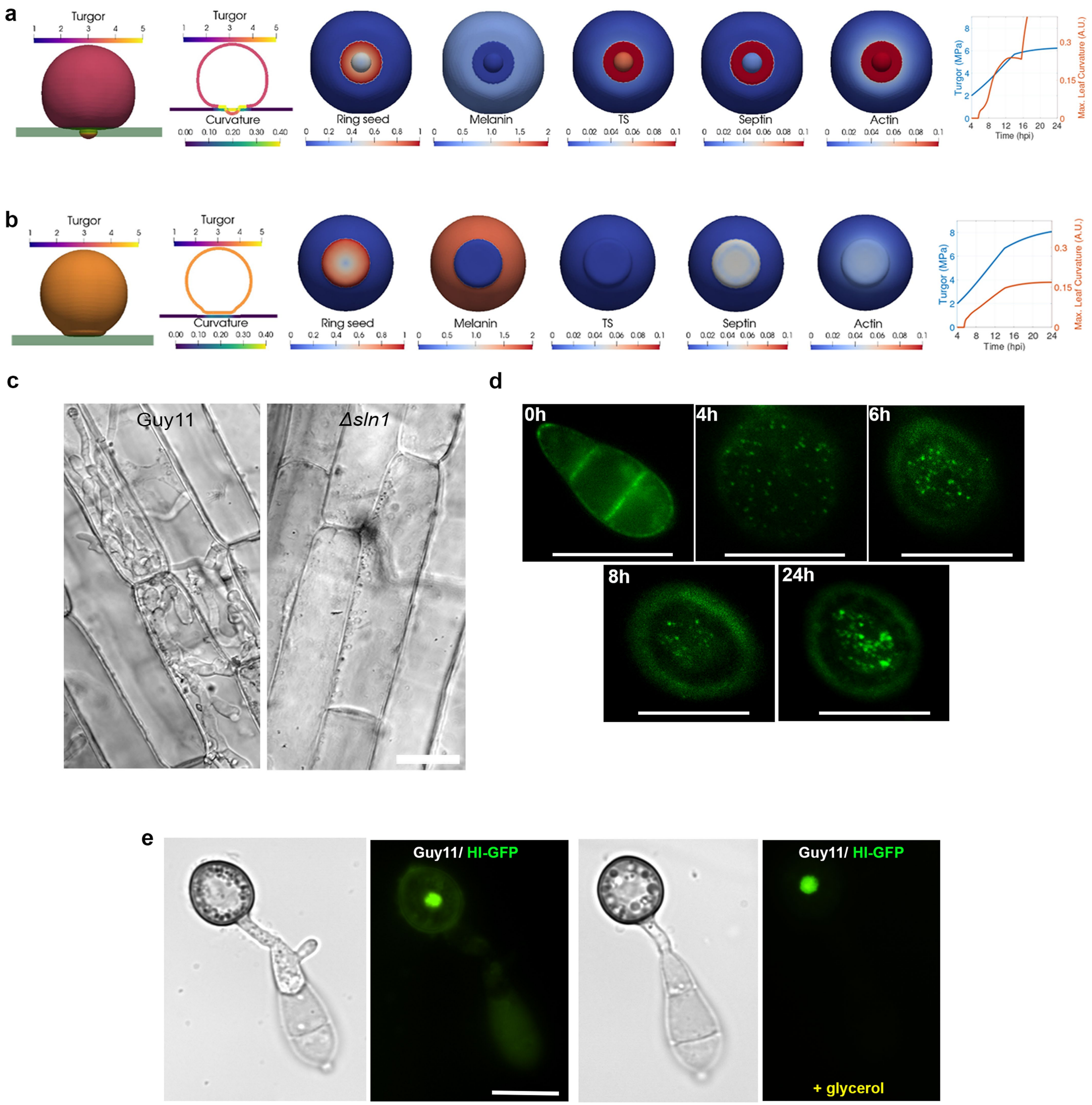
**Peer review information** *Nature* thanks Antonio Di Pietro, Nicholas Money and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



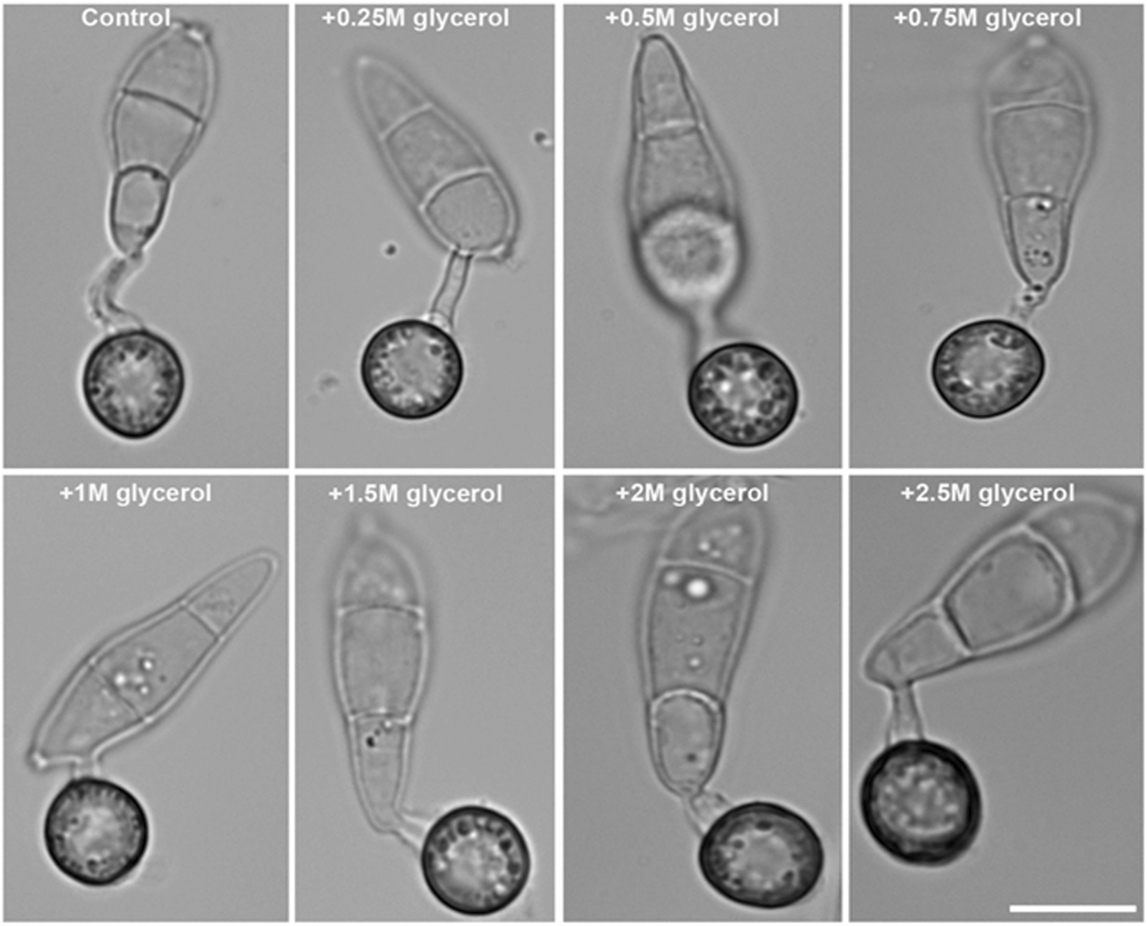
**Extended Data Fig. 1 | Application of glycerol to intact rice leaves does not cause cell collapse. a**, Micrographs showing epidermal strips of a transgenic rice line that expresses the plasma-membrane marker Lti6B–GFP, treated with either water or glycerol (5 M) for 24 h. Treatment with glycerol caused cell collapse. **b**, Intact leaves from 2-week-old Lti6B–GFP transgenic rice plants were inoculated with 30-μl drops of water or 5 M glycerol and incubated for 3 days. No plasmolysis was observed; that is, glycerol was unable to cause the collapse of cells in whole plants. **c**, Rice plants were treated with water or 5 M glycerol spray and incubated for 5 days ($n = 3$ independent replications of the experiment). Glycerol did not have any effect on the health of the rice plant or cause any wilting—confirming that no plasmolysis of rice cells from intact leaves occurs (as shown in **b**). Micrographs are representative of two independent replicates of the experiment. Scale bars, 20 μm (**a**); 5 μm (**b**).



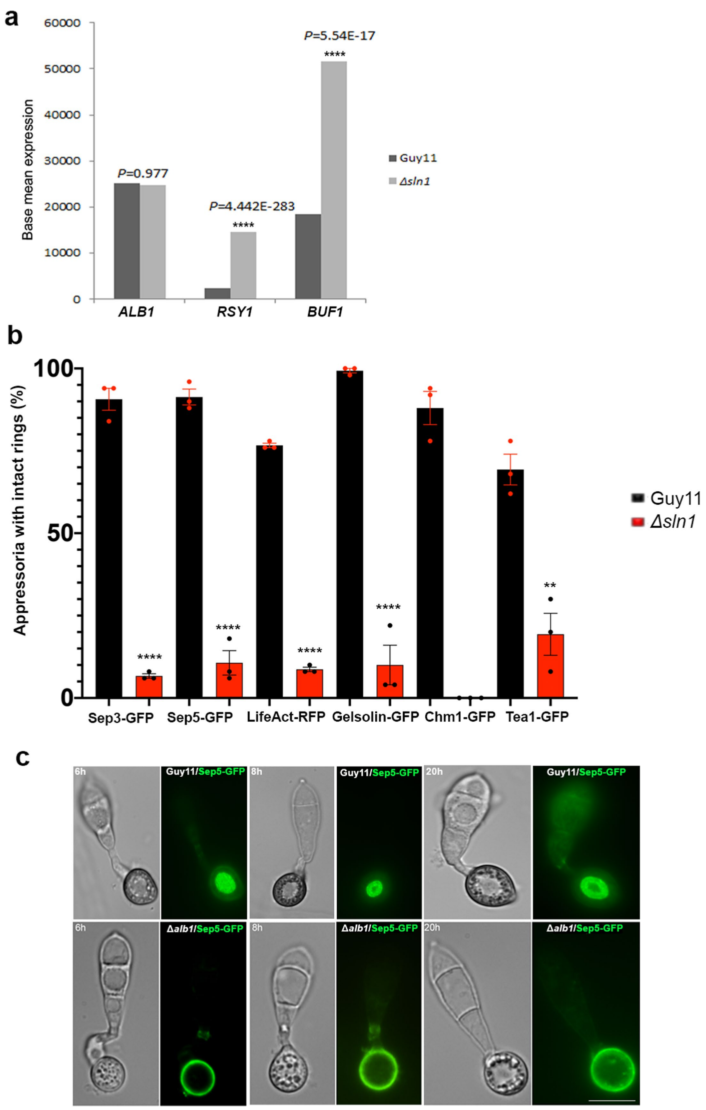
**Extended Data Fig. 2 | Septin organization is impaired by artificial lowering of turgor or inhibition of melanin biosynthesis. a**, Percentage of appressoria that have intact septin rings after treatment with 1.5 M glycerol or 100 μM tricyclazole. Treatments were applied between 0 and 20 h.p.i and quantified at 24 h.p.i. A window of effect could thus be defined for reaching the threshold of appressorium turgor (by 16–20 h.p.i.) and for completion of melanization (by 12 h.p.i. Data are mean ± s.d. for $n = 3$ independent biological replicates; 100 appressoria were counted per replicate. ****$P < 0.0001$,***$P < 0.001$, **$P < 0.01$ (two-tailed unpaired Student's *t*-test compared to untreated Guy11 control). **b**, Percentage of appressoria that have intact septin GTPase and F-actin rings in wild-type Guy11 and the melanin-deficient mutants Δ*alb1*, Δ*rsy1* and Δ*buf1* at 24 h.p.i. Data are mean ± s.d. for $n = 3$ independent biological replicates; 100 appressoria were counted per replicate.



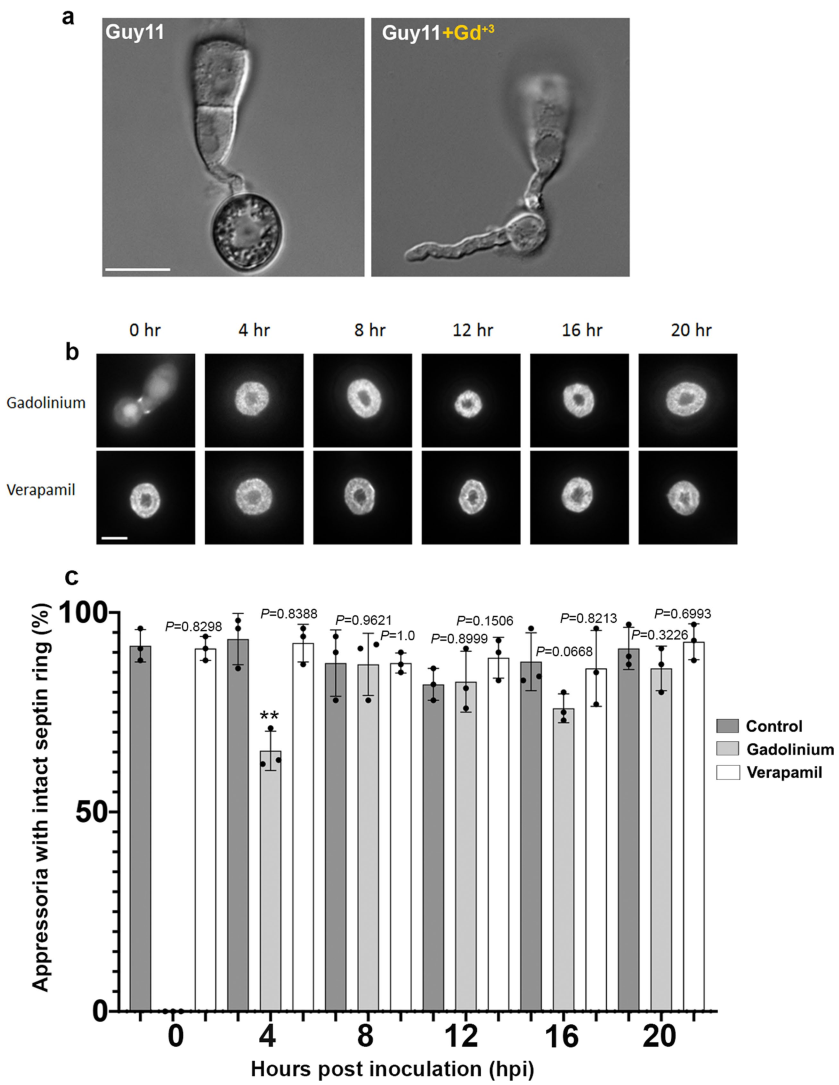
**Extended Data Fig. 3 | Characteristics of the Sln1 turgor-sensor kinase.** **a**, Graphical simulation of a mathematical model for appressorium function in *M. oryzae*. The model assumes that septins are recruited to the appressorium pore at a seeded ring structure that allows the recruitment of F-actin. Melanin is recruited to the appressorium dome in proportion to increasing turgor, and excluded from the pore. A turgor sensor (TS) is recruited to the pore to modulate melanization and turgor generation, while positively regulating septin recruitment and F-actin reorganization; this results in cuticle rupture (Supplementary Video 1). **b**, A mutant that lacks the turgor sensor generates excess appressorium turgor, recruits more melanin to the cell wall and prevents the recruitment of septin and F-actin to the pore; the cuticle is therefore not breached (Supplementary Video 2). **c**, Micrographs showing that the Δ*sln1* mutant is unable to invade and colonize rice tissue after 36 h.p.i. No invasive hyphae were visualized inside rice cells inoculated with Δ*sln1*. Images are representative of $n = 2$ independent biological replicates. Scale bar, 10 μm. **d**, Localization of Sln1–GFP in conidia and appressoria of *M. oryzae*. Conidia were collected from a *M. oryzae* Guy11 transformant that expresses a Sln1–GFP gene fusion, and inoculated on glass coverslips. Images were captured at 0, 4, 6, 8 and 24 h.p.i. Micrographs are representative of the distribution of Sln1–GFP at the indicated time points in $n = 3$ independent biological replications of the experiment. Scale bars, 10 μm. **e**, Epifluorescence micrographs showing that the cellular distribution of H1–GFP in appressoria at 24 h.p.i. is unaffected by exposure of appressoria to 1.5 M glycerol at 5 h.p.i. Images are representative of $n = 3$ independent biological replicates; 50 appressoria were counted per replicate. Scale bar, 10 μm.



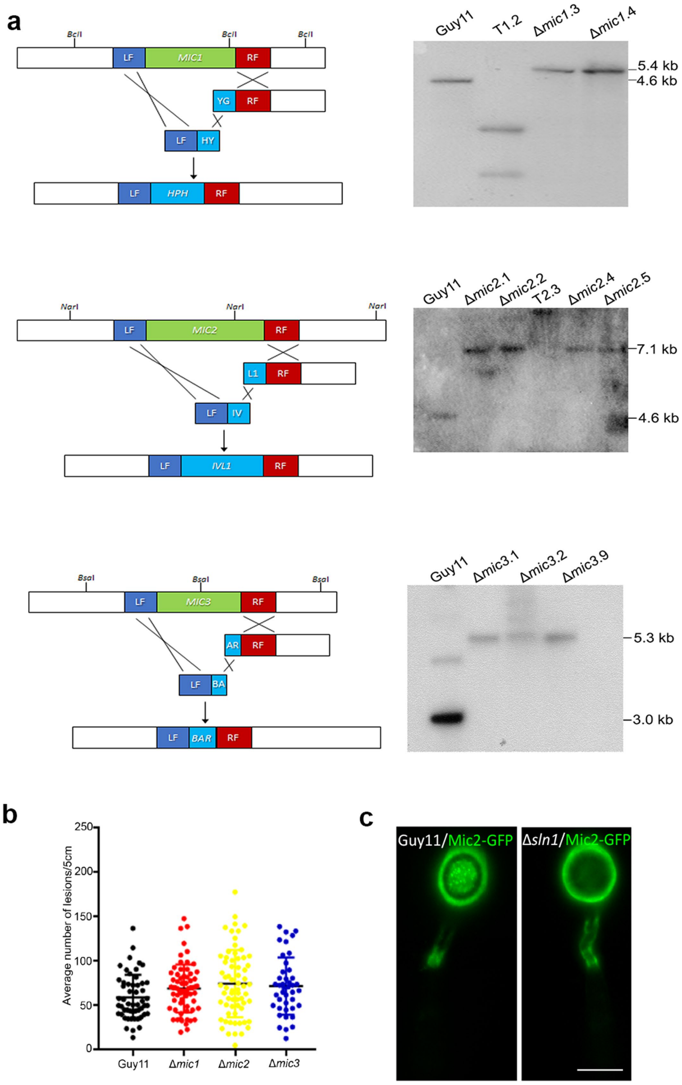
**Extended Data Fig. 4 | Deposition of melanin in appressoria increases in a dose-dependent manner after exposure to hyperosmotic stress.** Conidia were collected from the Guy11 strain and inoculated on glass coverslips. Glycerol solutions of different concentrations (ranging from 0.25 to 2.5 M) were applied 3–4 h.p.i. and the appressoria were imaged by bright-field microscopy at 24 h.p.i. to visualize the melanin layer in the appressorium. Artificially lowering turgor by application of hyperosmotic stress led to continual melanization of the appressorium—consistent with melanin biosynthesis and cell-wall deposition being turgor-dependent. Images are representative of $n = 2$ independent biological replications of the experiment. Scale bar, 10 μm.



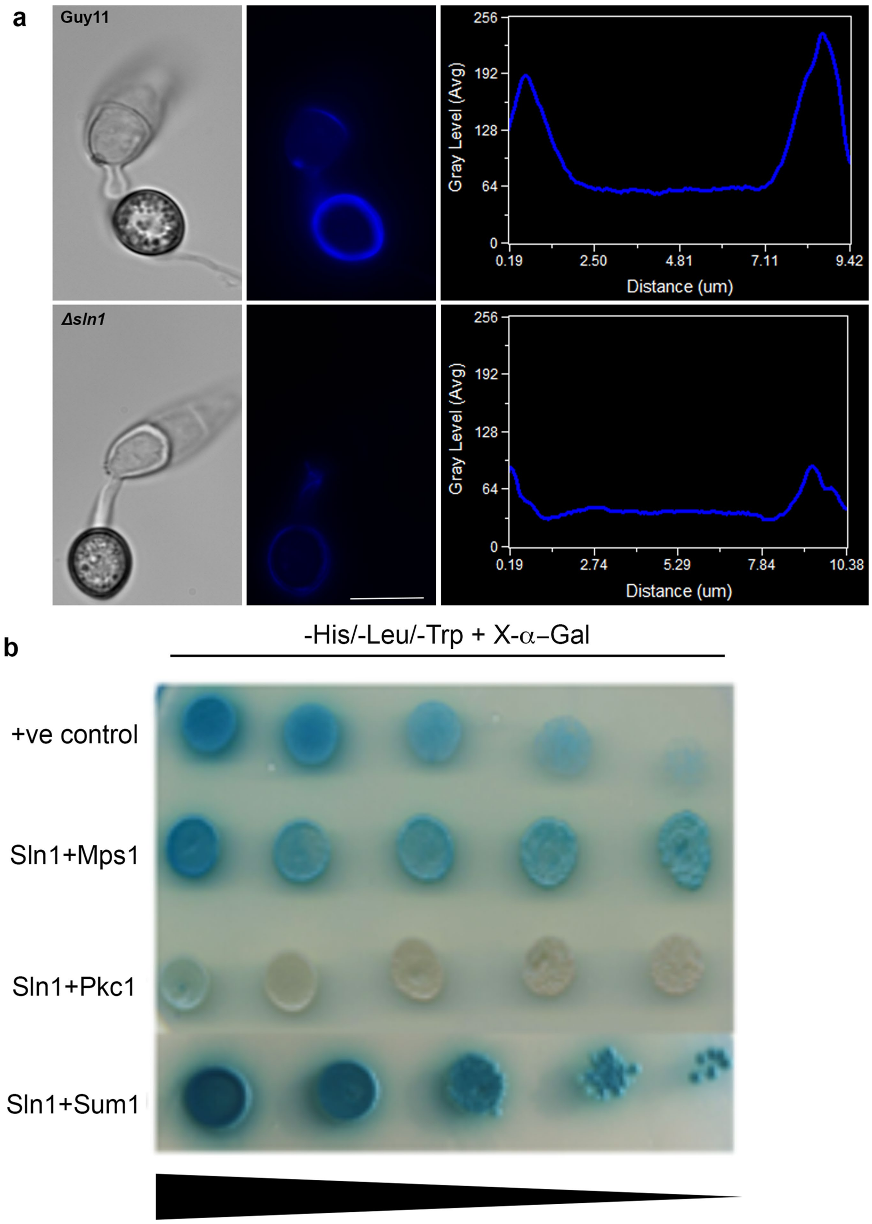
**Extended Data Fig. 5. | Melanin biosynthesis and cytoskeletal organization is affected in Δ*sln1* mutants. a**, Transcript abundance of genes that are involved in DHN-melanin biosynthesis in a Δ*sln1* mutant compared to Guy11 in appressoria at 16 h.p.i. Gene expression is represented as base mean expression from $n = 3$ three RNA-seq experiments. $****P < 0.0001$ (two-tailed unpaired Student's *t*-test). **b**, Sln1 is required for the septin-mediated reorganization of F-actin at the appressorium pore. Conidia were collected from Guy11 transformants that express Septin3–GFP, Septin5–GFP, LifeAct–RFP, gelsolin–GFP, Chm1–GFP and Tea1–GFP gene fusions, inoculated on glass coverslips and observed by epifluorescence microscopy at 24 h.p.i. The proportion of appressoria that have intact rings was recorded. Data are mean ± s.d. for $n = 3$ independent biological replicates; 50 appressoria were counted per replicate. $****P < 0.0001$, $**P < 0.01$ (two-tailed unpaired Student's *t*-test). **c**, Septin5–GFP is not recruited to the appressorium pore in a melanin-deficient Δ*alb1* mutant. Conidia were collected from Guy11 and Δ*alb1* transformants that express Sep5–GFP, inoculated on glass coverslips and observed by epifluorescence microscopy at 6, 8 and 20 h.p.i. The distribution of Sep5–GFP at the cell cortex or appressorium pore was recorded. Images are representative of $n = 3$ independent biological replicates. Scale bar, 10 μm.



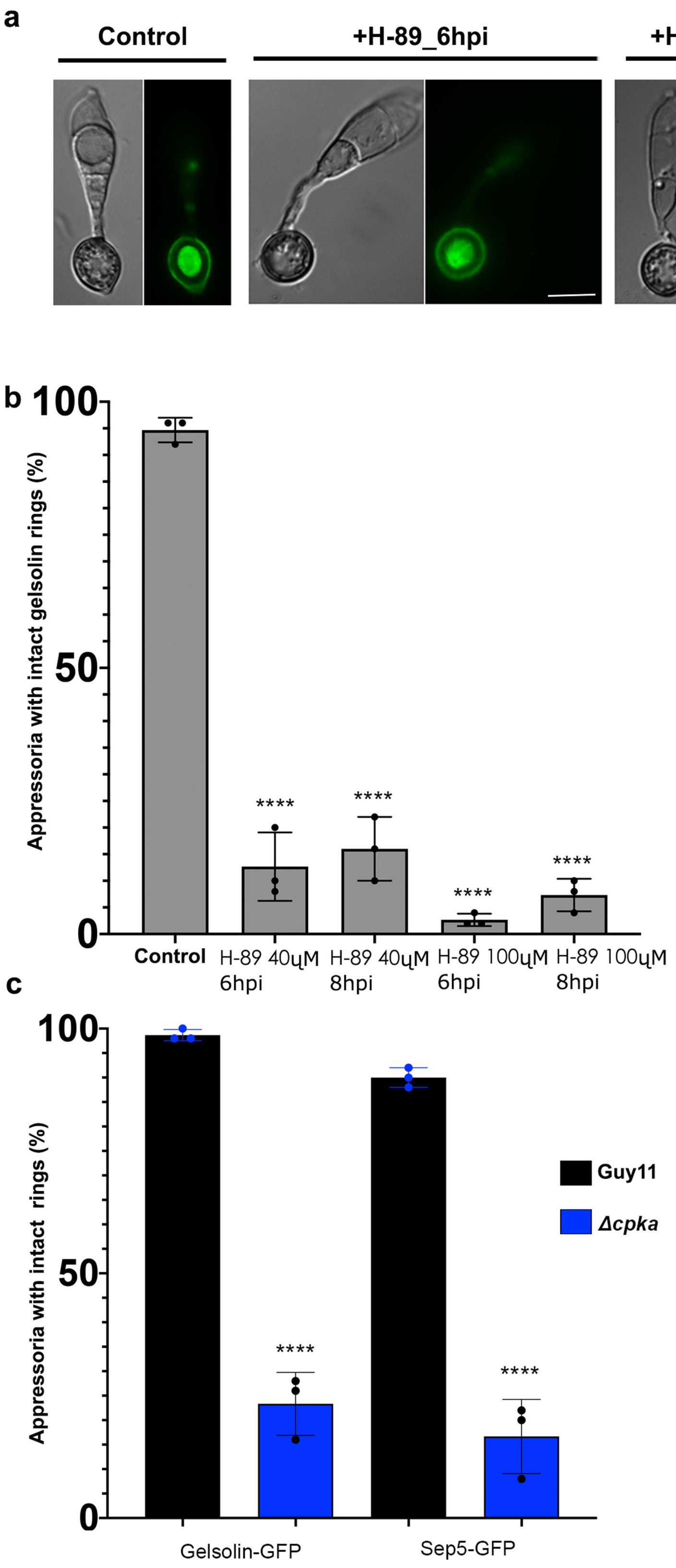
**Extended Data Fig. 6 | Mechanosensitive ion channels are required for appressorium formation but dispensable for septin-mediated cytoskeletal reorganization.** **a**, Conidia were collected from Guy11 and inoculated on glass coverslips in the presence and absence of gadolinium ($Gd^{+3}$). Appressoria were imaged by bright-field microscopy at 24 h.p.i. Addition of gadolinium disrupted appressorium formation. Images are representative of $n = 3$ independent biological replications of the experiment. **b**, Micrographs showing the cellular localization of Sep5–GFP at the appressorium pore of Guy11 after treatment with gadolinium or verapamil at 0–20 h.p.i., imaged at 24 h.p.i. Images are representative of $n = 3$ independent biological replicates. Scale bars, 10 μm (**a**, **b**). **c**, Percentage of appressoria that have intact septin rings after treatment with gadolinium and verapamil. Data are mean ± s.d. for $n = 3$ independent biological replicates; 100 appressoria were counted per replicate. $^{**}P < 0.01$ (two-tailed unpaired Student's *t*-test).



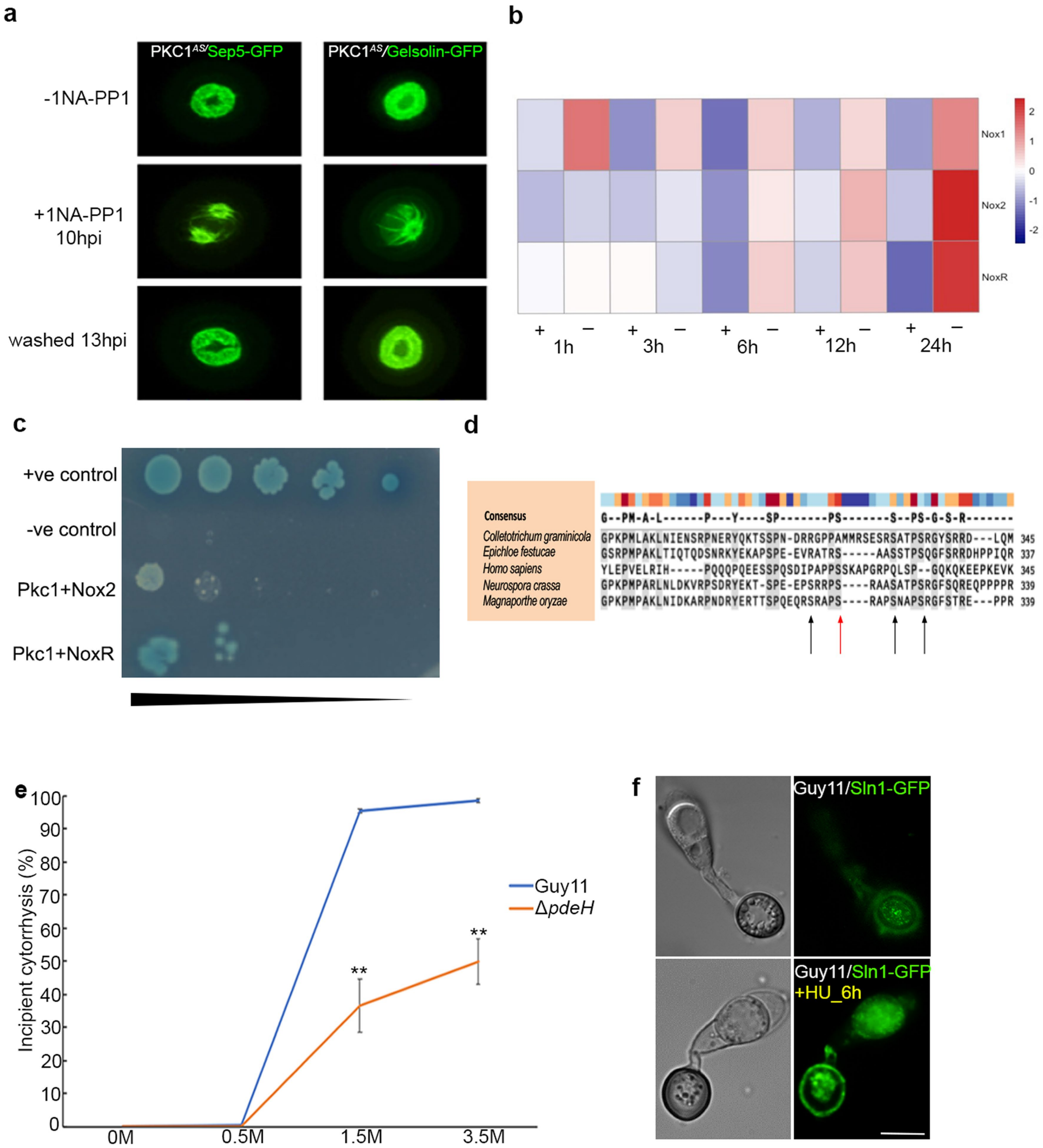
**Extended Data Fig. 7 | Mechanosensitive ion channels are dispensable for pathogenicity. a**, Schematic representation of the split-marker method that was used to generate targeted deletions of genes that encode mechanosensitive ion channels, LF denotes left flank and RF denotes right flank as regions flanking the open reading frame for the gene of interest. Primers are shown in Supplementary Table 1. Deletion mutants were identified by Southern blot analysis (for gel source data of the Δ*mic1*, Δ*mic2* and Δ*mic3* gene-deletion digests for Southern blotting, see Supplementary Fig. 1a, b and c, respectively). **b**, Dot plot showing the frequency of disease lesions observed in a 5-cm zone from each individual leaf harvested per strain (60 leaves were harvested per strain). Data are the mean and individual data points for $n = 2$ independent biological replicates. Each data point represents the number of lesions on an infected rice leaf. $P > 0.01$ (two-tailed, unpaired Student's *t*-test with Welch-correction, compared to a Guy11 control). **c**, Micrographs showing the cellular localization of Mic2–GFP at the appressorium pore of Guy11 and Δ*sln1* at 24 h.p.i. Images are representative of $n = 3$ independent biological replicates. Scale bar, 10 μm.



**Extended Data Fig. 8 | Chitin deposition in the appressorium cell wall is impaired in a Δ*sln1* mutant. a**, Conidia were collected from Guy11 and the Δ*sln1* mutant and inoculated on glass coverslips to form appressoria. At 24 h.p.i., appressoria were stained with 50 μM calcofluor white for 5 min in the dark, washed and images captured by epifluorescence microscopy. Line-scan graphs represent calcofluor white fluorescence in a transverse section of an individual appressorium. Images are representative of $n = 3$ independent biological replicates. Scale bar, 10 μm. **b**, The Sln1 kinase interacts with Sum1, Pkc1 and Mps1 in a yeast two-hybrid assay. Simultaneous co-transformation of pGAD-Sln1 (prey vector) with pGBK-Mps1, pGBK-PKC and pGBK-Sum1 (bait vectors) and pGBKT7-53 and pGADT7-T (positive-control vectors) into the Y2H Gold strain resulted in the activation of three reporter genes and growth on medium-stringency medium (−Ade, −Leu, –Trp, +X-α-gal). Co-transformation also activates the expression of *MEL1*, which results in the secretion of α–galactosidase and the hydrolysis of X-α-gal in the medium, turning the yeast colonies blue. Images are representative of $n = 2$ biological replications of the experiment.



**Extended Data Fig. 9 | The PKA inhibitor H-89 disrupts gelsolin ring assembly. a**, Micrographs showing the cellular localization of gelsolin–GFP at the appressorium pore of Guy11 after treatment with the PKA inhibitor H-89 at 6 h.p.i. and 8 h.p.i., imaged at 24 h.p.i. Images are representative of $n = 3$ independent biological replications of the experiment. Scale bar, 10 μm. **b**, Percentage of appressoria that have intact gelsolin rings after treatment with H-89. Data are mean ± s.d. for $n = 3$ independent biological replicates; 50 appressoria were counted per replicate. $^{****}P < 0.0001$ (two-tailed unpaired Student's *t*-test). **c**, Percentage of appressoria that have intact gelsolin and septin rings in Guy11 and the Δ*cpka* mutant. Rings were routinely smaller in Δ*cpka* than in Guy11, indicating the reduced diameter of the appressorium in the Δ*cpka* mutant. Data are mean ± s.d. for $n = 3$ independent biological replicates; 50 appressoria were counted per replicate. $^{****}P < 0.0001$ (two-tailed unpaired Student's *t*-test).



**Extended Data Fig. 10 | Interplay between the cell integrity and cAMP-dependent protein kinase A pathways in turgor sensing by the rice blast fungus. a**, Inhibition of Pkc1 activity with 1NA-PP1 can be reversed to restore septin and gelsolin ring formation. Micrographs showing the cellular localization of Sep5–GFP and gelsolin–GFP at the appressorium pore following Pkc1 blocking at 10 h.p.i. and releasing at 13 h.p.i. Appressoria were imaged at 24 h.p.i. Images are representative of $n = 2$ replications of the experiment. **b**, Heat map showing the expression of *NOX1*, *NOX2* and *NOXR* in an RNA-seq analysis of the $PKC1^{AS}$-mutant. Mycelium was grown in CM shake cultures for 48 h in the presence or absence of 500 nM 1NA-PP1 at 1, 3, 6, 12 or 24 h.p.i. ($n = 3$ biological replications of the experiment). The full RNA-seq dataset from this study can be found at the Gene Expression Omnibus (GEO) under accession number GSE70308. **c**, Pkc1 interacts with Nox2 and NoxR in a yeast two-hybrid assay. Simultaneous co-transformation of pGBK-PKC (bait vector) and pGAD-Nox2 and pGAD-NoxR (prey vectors) into the Y2H Gold strain resulted in the activation of three reporter genes and growth on medium-stringency medium for Pkc1 and NoxR (−His, −Leu, –Trp, +X-α-gal), and on high-stringency medium for Pkc1 and Nox2 (–His, −Leu, −Trp, −Ade, +X-α-gal). Co-transformation also activates the expression of *MEL1*, which results in the secretion of α–galactosidase and the hydrolysis of X-α-gal in the medium, turning the yeast colonies blue. Images are representative of $n = 3$ independent biological replications of the experiment. **d**, Alignment of a region of the predicted amino acid sequence of NoxR using Muscle[40]. Sequence conservation is shaded in grey, with a consensus threshold of 75%. The predicted Pkc1 phosphorylation site is marked with a red arrow and is highly conserved; black arrows indicate other potential Pkc1 phosphorylation targets based on the motif S*APS. **e**, Δ*pdeH* mutants generate excess appressorium turgor. Percentage of Guy11 and Δ*pdeH*-mutant appressoria that undergo incipient cytorrhysis after exposure to glycerol solutions of 0–3.5 M. Data are mean ± s.e.m. for $n = 3$ independent biological replicates; 50 appressoria were counted per replicate. $^{**}P < 0.01$ (two-tailed unpaired Student's *t*-test). **f**, Cellular localization of Sln1–GFP in appressorium pores of Guy11, with or without hydroxyurea (HU) added at 6 h.p.i and imaged at 24 h.p.i. Images are representative of $n = 3$ independent biological replicates. Scale bars, 10 μm.

nature research

Corresponding author(s): 2017-10-13477C Nicholas J Talbot

Last updated by author(s): Jul 24, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Images were captured by using a Photometrics CoolSNAP HQ2 camera (Roper Scientific), under control of Metamorph software version 7.7.4 (MDS Analytical Technologies). For mass spectromtery, MS/MS peak lists were exported using Discoverer version 2.2 (Thermo Scientific). |
| Data analysis | Proteomic data analysis used Mascot v 2.3 (Matrix Science) verified with Scaffold (Proteome Software). For RNA-seq analysis TopHat v2.1 software was used and analysis of data performed using DESeq. Quantitative data was routinely analyzed using GraphPad Prism version 8 and Microsoft Excel for Mac version 16.16.12. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

*Provide your data availability statement here.*

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample sizes were as large as practicable for observations of appressorium function and virulence studies, based on previous studies where estimates have provide statistically significant findings and prior experience. No statistical methods were used to pre-determine sample size. |
| Data exclusions | No data were excluded |
| Replication | All experiments were subject to three independent biological replications unless otherwise stated. Technical replications were also carried out as stated in the text. |
| Randomization | All experimental observations were carried out without pre-selection of groups. no other form of randomization was relevant to the study. |
| Blinding | Blind testing was not routinely carried out in the study as it was not relevant to most of the experiments carried out, as these were based on molecular genetic and biochemical characterization of the Sln1 turgor sensing complex, requiring knowledge of samples to be processed. Blind testing of virulence phenotypes was carried out as an independent check of phenotypes, wherever possible, such as the information provided in Figure 1b. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

https://doi.org/10.1038/s41586-019-1548-x

# Cleavage of RIPK1 by caspase-8 is crucial for limiting apoptosis and necroptosis

Kim Newton[1]*, Katherine E. Wickliffe[1], Debra L. Dugger[1], Allie Maltzman[1], Merone Roose-Girma[2], Monika Dohse[3], László Kőműves[3], Joshua D. Webster[3] & Vishva M. Dixit[1]*

**The aspartate-specific cysteine protease caspase-8 suppresses necroptotic cell death mediated by RIPK3 and MLKL. Indeed, mice that lack caspase-8 die in a RIPK3- and MLKL-dependent manner during embryogenesis[1–3]. In humans, caspase-8 deficiency is associated with immunodeficiency[4] or very early onset inflammatory bowel disease[5]. The substrates that are cleaved by caspase-8 to prevent necroptosis in vivo have not been defined. Here we show that knock-in mice that express catalytically inactive caspase-8(C362A) die as embryos owing to MLKL-dependent necroptosis, similar to caspase-8-deficient mice. Thus, caspase-8 must cleave itself, other proteins or both to inhibit necroptosis. Mice that express caspase-8(D212A/D218A/D225A/D387A), which cannot cleave itself, were viable, as were mice that express c-FLIP or CYLD proteins that had been mutated to prevent cleavage by caspase-8. By contrast, mice that express RIPK1(D325A), in which the caspase-8 cleavage site Asp325 had been mutated, died mid-gestation. Embryonic lethality was prevented by inactivation of RIPK1, loss of TNFR1, or loss of both MLKL and the caspase-8 adaptor FADD, but not by loss of MLKL alone. Thus, RIPK1(D325A) appears to trigger cell death mediated by TNF, the kinase activity of RIPK1 and FADD–caspase-8. Accordingly, dying endothelial cells that contain cleaved caspase-3 were abnormally abundant in yolk sacs of *Ripk1*$^{D325A/D325A}$ embryos. Heterozygous *Ripk1*$^{D325A/+}$ cells and mice were viable, but were also more susceptible to TNF-induced cell death than were wild-type cells or mice. Our data show that Asp325 of RIPK1 is essential for limiting aberrant cell death in response to TNF, consistent with the idea that cleavage of RIPK1 by caspase-8 is a mechanism for dismantling death-inducing complexes.**

Caspase-8 is essential for apoptosis triggered by death receptors[6], but it also prevents death receptors, the Toll-like receptors TLR3 and TLR4, and T-cell receptors from inducing necroptosis[1–3,7,8]. The pro-survival function of caspase-8 is essential for mouse development because *Casp8*$^{-/-}$ mice die around embryonic day 11 (E11)[6] from unchecked necroptosis[1–3]. Caspase-8 may form heterodimers with c-FLIP to cleave and inactivate proteins that promote necroptosis[2]. To confirm that the catalytic activity of caspase-8 is required for mouse development, we generated *Casp8*$^{C362A/C362A}$ knock-in mice expressing catalytically inactive caspase-8(C362A) (Extended Data Fig. 1a). Similar to *Casp8*$^{-/-}$ mice[6] and mice lacking caspase-8 only in endothelial cells[9], *Casp8*$^{C362A/C362A}$ embryos had abnormal yolk sac vasculature (Fig. 1a) and died between E10.5 and E12.5 (Table 1). At E10.5, *Casp8*$^{C362A/C362A}$ embryos contained autophosphorylated RIPK3 (Fig. 1b), a hallmark of necroptosis that is also seen in *Casp8*$^{-/-}$ embryos[10]. Accordingly, MLKL deficiency allowed *Casp8*$^{C362A/C362A}$ embryos to survive to birth (Fig. 1c, Table 1). As expected, *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ mouse embryo fibroblasts (MEFs) did not cleave a caspase-8 substrate after treatment with cycloheximide and TNF (Extended Data Fig. 1b), and were as resistant as *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$ MEFs to apoptosis induced by cycloheximide plus TNF, FasL or TRAIL (Extended Data Fig. 1c). Expression of caspase-8(C362A) was slightly less than that of wild-type caspase-8, but autophosphorylated RIPK3 and autophosphorylated RIPK1 were more abundant in *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ MEFs than in *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$ MEFs (Extended Data Fig. 1d). Moreover, in contrast to *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$ mice (which are viable[3]), *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ mice died soon after birth (Table 1). Therefore, the caspase-8(C362A) scaffold must cause perinatal lethality independent of necroptosis. At E18.5, *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ and *Mlkl*$^{-/-}$ littermates were similar in weight (Extended Data Fig. 1e). The only morphological abnormality in *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ embryos was villous atrophy of the small intestine (Fig. 1d). This perinatal phenotype will be characterized elsewhere, but the survival of *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ mice to birth establishes that the proteolytic activity of caspase-8 limits necroptosis during embryogenesis.

Caspase-8 undergoes autoprocessing[11] and its other proposed substrates include c-FLIP[12], CYLD[13], RIPK1[14], and RIPK3[15]. If cleavage of one of these proteins is required to prevent necroptosis during development, then mutation of the Asp residues that are cleaved by caspase-8 should mimic inactivation of caspase-8 and be lethal at mid-gestation. Transgenic expression of caspase-8(D387A), which cannot be cleaved between the large and small catalytic subunits, rescues the lethality of *Casp8*$^{-/-}$ mice[16], which indicates that the Asp387 autoprocessing site is not required for the pro-survival function of caspase-8. We confirmed that *Casp8*$^{1\times DA/1\times DA}$ knock-in mice expressing caspase-8(D387A) were viable (Extended Data Fig. 1e, f), but caspase-8(D387A) in thymocytes was still processed between its pro-domain and large catalytic subunit after treatment with FasL (Extended Data Fig. 1g). Therefore, we also mutated Asp212, Asp218 and Asp225 in this region (Extended Data Fig. 1a). *Casp8*$^{4\times DA/4\times DA}$ mice expressing caspase-8(D212A/D218A/D225A/D387A) were viable (Extended Data Fig. 1h) and we detected no processing of the mutant caspase-8 in response to FasL (Extended Data Fig. 1g). *Casp8*$^{4\times DA/4\times DA}$ and *Casp8*$^{1\times DA/1\times DA}$ thymocytes were killed by FasL, albeit not as rapidly as were wild-type thymocytes (Extended Data Fig. 1i), as previously found[16]. Accordingly, lysates from FasL-treated *Casp8*$^{4\times DA/4\times DA}$ thymocytes exhibited delayed cleavage of a caspase-8 substrate when compared with lysates from wild-type cells (Extended Data Fig. 1j). Given that interdomain cleavage of caspase-8 is essential for the proteolytic activity of caspase-8 homodimers, but is dispensable for the activity of c-FLIP–caspase-8 heterodimers[12], it may be that c-FLIP–caspase-8 heterodimers mediate FasL-induced apoptosis in *Casp8*$^{4\times DA/4\times DA}$ cells. Consistent with the relative preservation of FasL-induced apoptosis, neither *Casp8*$^{1\times DA/1\times DA}$ nor *Casp8*$^{4\times DA/4\times DA}$ mice developed the lymphadenopathy and splenomegaly that occurs in Fas-deficient mice[17] (Extended Data Fig. 1k–n). Thus, autoprocessing of caspase-8 is not required for preventing necroptosis during development or for FasL-induced apoptosis.

*Cflar*$^{2\times DA/2\times DA}$ mice expressing c-FLIP(D371A/D377A), which cannot be cleaved by caspase-8 (Extended Data Fig. 2a), were also viable (Extended Data Fig. 2b). *Cflar*$^{2\times DA/2\times DA}$ thymocytes exhibited slightly slower FasL-induced apoptosis than did wild-type thymocytes

---

[1]Department of Physiological Chemistry, Genentech, South San Francisco, CA, USA. [2]Department of Molecular Biology, Genentech, South San Francisco, CA, USA. [3]Department of Pathology, Genentech, South San Francisco, CA, USA. *e-mail: knewton@gene.com; dixit@gene.com

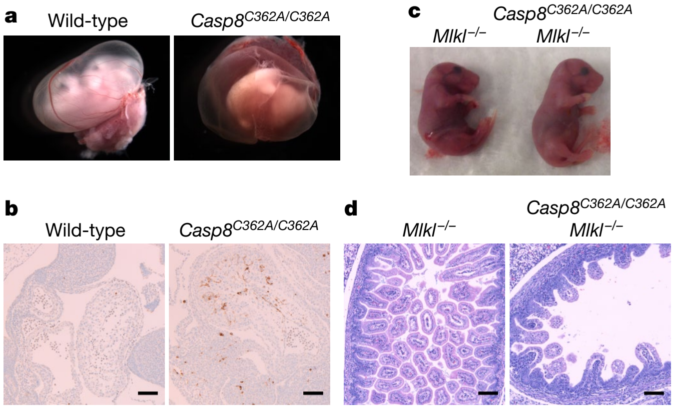


**Fig. 1 | Catalytic activity of caspase-8 prevents necroptosis during mouse embryogenesis.** **a**, E12.5 embryos representative of two wild-type and three *Casp8*$^{C362A/C362A}$ embryos. **b**, E10.5 embryo sections immunolabelled for phosphorylated (p)-RIPK3(T231/S232) (brown). Results representative of three wild-type and three *Casp8*$^{C362A/C362A}$ embryos. Scale bar, 100 μm. **c**, E18.5 embryos representative of 19 *Mlkl*$^{-/-}$ and 22 *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ embryos. **d**, E18.5 embryo sections stained with haematoxylin and eosin. Results representative of five *Mlkl*$^{-/-}$ and five *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$ embryos. Scale bar, 100 μm.

(Extended Data Fig. 2c). However, as in *Casp8*$^{4\times DA/4\times DA}$ mice, this defect did not cause lymphadenopathy (Extended Data Fig. 2d). Thus, c-FLIP is not an essential substrate of caspase-8 during embryogenesis. *Cyld*$^{D215A/D215A}$ mice expressing CYLD(D215A) (Extended Data Fig. 2e, f) were also viable (Extended Data Fig. 2g, h), indicating that cleavage by caspase-8 after CYLD Asp215 is not required to prevent necroptosis. CYLD(D215A) still underwent caspase-dependent cleavage in thymocytes treated with FasL (Extended Data Fig. 2f). The 24-kDa N-terminal fragment associated with cleavage at Asp215 was eliminated, but a larger fragment appeared in its place.

*Ripk1*$^{D325A/D325A}$ mice expressing RIPK1(D325A), which cannot be cleaved by caspase-8 (Extended Data Fig. 3a), were not viable (Table 2), and had abnormal yolk sac vasculature at E10.5 (Fig. 2a). Similar lethality was reported for an independent *Ripk1*$^{D325A/D325A}$ mouse strain[18]. The death of the *Ripk1*$^{D325A/D325A}$ embryos was not caused by necroptosis, because neither *Ripk1*$^{D325A/D325A}$ *Ripk3*$^{-/-}$ nor *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ mice were viable (Table 2). Indeed, *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ embryos also had abnormal yolk sac vasculature at E10.5 (Fig. 2a). As previously reported[18], *Ripk1*$^{D325A/D325A}$ *Ripk3*$^{-/-}$ embryos looked normal at E12.5 (Extended Data Fig. 3b). Therefore, RIPK3 must contribute to the lethality of *Ripk1*$^{D325A/D325A}$ embryos at E10.5 in a necroptosis-independent manner. Immunolabelling for cleaved caspase-3 suggested that RIPK1(D325A) caused apoptosis in the yolk sac vasculature (Fig. 2b). RIPK1, similar to caspase-8 and RIPK3, is expressed in endothelial cells and other cell types at E10.5[19] (Extended Data Fig. 3c). Embryonic lethality required the enzymatic activity of RIPK1 and the presence of TNFR1, FADD and caspase-8, because *Ripk1*$^{D138N,D325A/D138N,D325A}$, *Ripk1*$^{D325A/D325A}$ *Tnfr1*$^{-/-}$ (*Tnfr1* is also known as *Tnfrsf1a*), *Ripk1*$^{D325A/D325A}$ *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$, and *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice survived to birth (Table 2). *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice were viable (Fig. 2c), but developed lymphadenopathy and splenomegaly similar to *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice (Extended Data Fig. 3d–f), whereas *Ripk1*$^{D138N,D325A/D138N,D325A}$ and *Ripk1*$^{D325A/D325A}$ *Tnfr1*$^{-/-}$ mice were runted (Extended Data Fig. 4a, b), exhibited granulocytic infiltrates and inflammation in multiple tissues (Extended Data Fig. 4c, d), and to date have not survived to weaning. Thus, in neonates, RIPK1(D325A) must promote cell death independent of TNFR1 and its own kinase activity. Consistent with caspase-8-dependent cleavage of RIPK1 inactivating both the scaffold and enzymatic death-promoting functions of RIPK1, catalytically inactive RIPK1(D138N) delayed lethality in *Casp8*$^{-/-}$ embryos only to around E14 (Extended Data Fig. 4e–g), whereas *Casp8*$^{-/-}$ *Ripk1*$^{-/-}$ mice survive to birth[20–22]. As expected, RIPK1(D325A) did not delay the death of *Casp8*$^{-/-}$ embryos (Extended Data Fig. 4h).

**Table 1 | Offspring numbers from intercrossing *Casp8*$^{C362A/+}$ mice**

| Genetic background | Age | *Casp8*+/+ | *Casp8*$^{C362A/+}$ | *Casp8*$^{C362A/C362A}$ |
|---|---|---|---|---|
| Wild-type | E10.5 | 6 | 7 | 5 |
| Wild-type | E12.5 | 2 | 5 | 3[a] |
| Wild-type | P4–P7[a] | 101 | 189 | 0 |
| *Mlkl*–/– | E18.5 | 19 | 37 | 22 |
| *Mlkl*–/– | P0 | 6 | 15 | 3 + 4[b] |
| *Mlkl*–/– | P4–P7 | 69 | 109 | 0 |

[a] P, postnatal day.
[b] Resorbing or dead.

**Table 2 | P4–P7 offspring numbers from intercrossing *Ripk1*$^{D325A/+}$ mice**

| Genetic background | *Ripk1*$^{+/+}$ | *Ripk1*$^{D325A/+}$ | *Ripk1*$^{D325A/D325A}$ |
|---|---|---|---|
| Wild-type | 21 | 38 | 0 |
| *Mlkl*$^{-/-}$ | 55 | 121 | 0 |
| *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ | 17 | 35 | 26 |
| *Ripk3*$^{-/-}$ | 34 | 62 | 0 |
| *Ripk3*$^{-/-}$ *Casp8*$^{-/-}$ | 3 | 15 | 5 |
| *Ripk1*$^{D138N/D138N}$ | 27 | 62 | 12 |
| *Tnfr1*$^{-/-}$ | 20 | 38 | 4 |

RIPK1 contributes to TNF-induced activation of NF-κB and MAPK in MEFs[23], but *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ MEFs phosphorylated IκBα, RelA, JNK, p38 MAPK and ERK normally in response to TNF (Extended Data Fig. 4i). Thus, defects in TNF-induced NF-κB and MAPK signalling are unlikely to cause the FADD–caspase-8-dependent cell death that is triggered by RIPK1(D325A) during embryogenesis.

*Ripk1*$^{D325A/D325A}$ MEFs were viable, and their expression of RIPK1(D325A) was comparable to that of RIPK1 in wild-type MEFs (Fig. 2d). However, *Ripk1*$^{D325A/D325A}$ MEFs treated with TNF showed increased autophosphorylation of RIPK1 and RIPK3, increased cleavage of caspase-8, caspase-3, and c-FLIP (Fig. 2d), and increased cell death (Fig. 2e) when compared to wild-type MEFs. The pan-caspase inhibitor carbobenzoxy-valyl-alanyl-aspartyl-[O-methyl]-fluoromethylketone (Z-VAD-FMK) largely blocked this cell death, despite the cells containing active RIPK3 (Fig. 2e). Wild-type MEFs treated with TNF and Z-VAD-FMK, untreated *Ripk1*$^{-/-}$ MEFs and untreated MEFs expressing RIPK1 with a mutant RIP homotypic interaction motif also activate RIPK3 with minimal necroptosis[23], perhaps because primary MEFs counter the effects of phosphorylated MLKL[24]. In contrast to the embryonic lethality of *Ripk1*$^{D325A/D325A}$ mice (which required catalytically active RIPK1), the death of *Ripk1*$^{D325A/D325A}$ MEFs exposed to TNF was not blocked by inhibition of RIPK1 with necrostatin-1 (Nec-1) (Fig. 2e). However, TNF was less effective at killing *Ripk1*$^{D325A/D325A}$ *Ripk3*$^{-/-}$ MEFs than *Ripk1*$^{D325A/D325A}$ MEFs (Extended Data Fig. 5a, b). The RIPK3 scaffold appears to be important because inhibition of RIPK3 with GSK′843[25] offered no benefit (Extended Data Fig. 5b). A kinase-independent role for RIPK3 in caspase-8-dependent apoptosis has also been reported in immortalized MEFs killed with TNF plus an inhibitor of TAK1, or TNF plus an IAP antagonist[26].

If RIPK1(D325A) enhances the stability of death-inducing complexes because it is not cleaved by caspase-8, then even one allele of RIPK1(D325A) might compromise cell survival. Indeed, *Ripk1*$^{D325A/+}$ cells were more sensitive than wild-type cells to apoptosis induced by either TNF plus cycloheximide (Fig. 3) or TNF plus TAK1 inhibitor (Extended Data Fig. 5c, d). *Ripk1*$^{D325A/+}$ bone-marrow-derived macrophages were also more sensitive than wild-type bone-marrow-derived macrophages to necroptosis induced by TNF plus Z-VAD-FMK (Extended Data Fig. 5e, f). Although RIPK1(D325A) enhanced

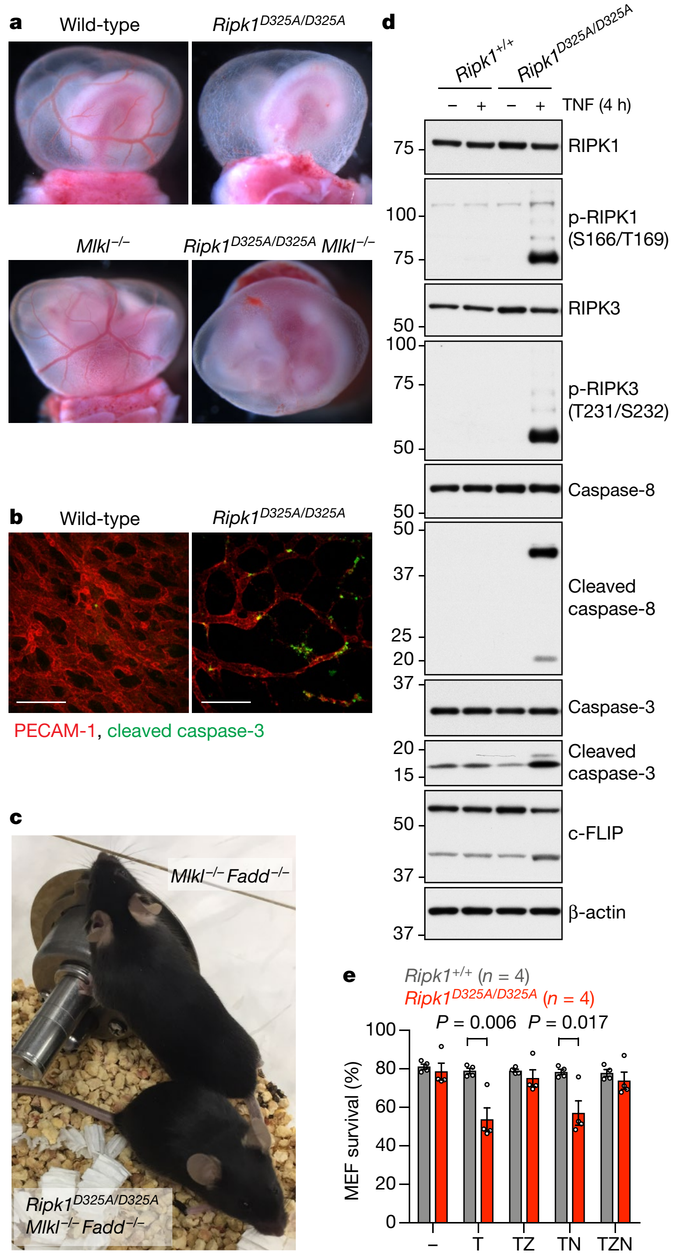


**Fig. 2 | RIPK1(D325A) triggers FADD-dependent cell death. a**, E10.5 embryos representative of five wild-type, four *Ripk1*$^{D325A/D325A}$, six *Mlkl*$^{-/-}$ and nine *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ embryos. **b**, E10.5 yolk sacs immunolabelled for PECAM-1 (red) and cleaved caspase-3 (green). Scale bar, 100 μm. Results representative of two wild-type and three *Ripk1*$^{D325A/D325A}$ embryos. **c**, Littermate males aged eight weeks. Results representative of 25 *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ and 21 *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice. **d**, Western blots of primary MEFs. β-Actin loading control performed after cleaved caspase-8. For gel source data, see Supplementary Fig. 1. Results representative of two independent experiments. **e**, Graph shows the percentage of viable MEFs after 24 h (−, untreated; T, TNF; Z, Z-VAD-FMK; N, Nec-1). Circles, cells from different embryos. Bars, mean ± s.e.m. *P* values (unpaired, two-sided *t*-test) shown if $P < 0.05$.

autophosphorylation of RIPK1 and RIPK3 in MEFs treated with TNF plus cycloheximide (Fig. 3a), only Z-VAD-FMK afforded *Ripk1*$^{D325A/+}$ MEFs some protection from this stimulus, whereas Nec-1 offered no benefit (Fig. 3b). Nec-1 reduced the death of only *Ripk1*$^{D325A/+}$ cells treated with TNF plus cycloheximide plus Z-VAD-FMK (Fig. 3a), TNF plus TAK1 inhibitor (Extended Data Fig. 5c, d) or TNF plus Z-VAD-FMK (Extended Data Fig. 5e), consistent with the kinase activity of RIPK1 being critical for these forms of TNF-induced cell death[26,27].

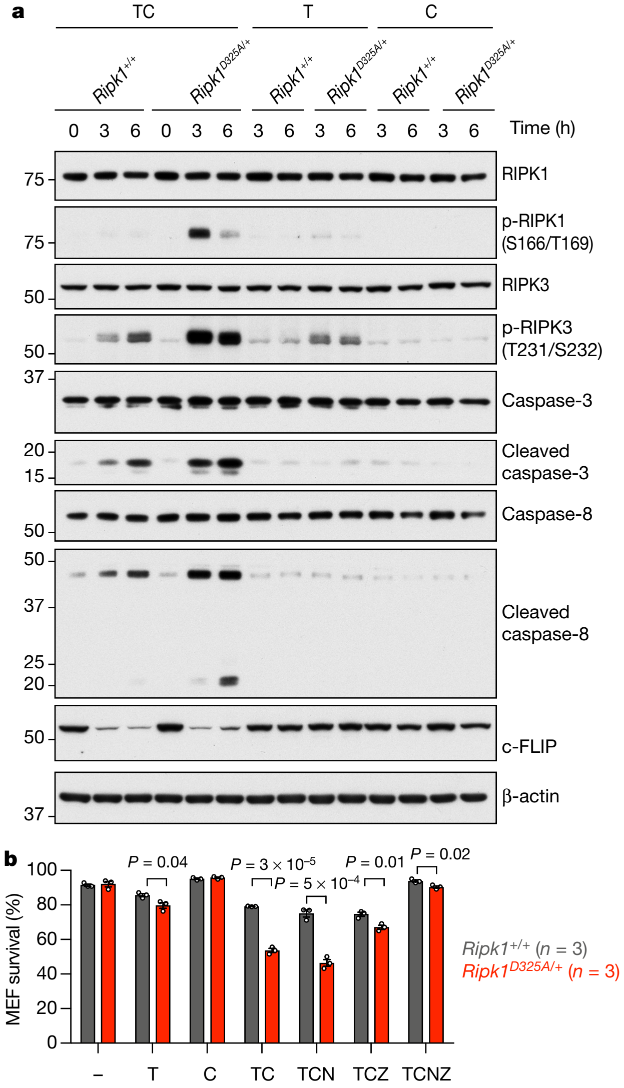


**Fig. 3 | RIPK1(D325A) heterozygosity sensitizes cells to TNF toxicity. a**, Western blots of primary MEFs. β-Actin loading control performed after caspase-8. For gel source data, see Supplementary Fig. 1. Results representative of two independent experiments. **b**, Graph shows the percentage of viable MEFs after 19 h (−, untreated; T, TNF; C, cycloheximide; Z, Z-VAD-FMK; N, Nec-1). Circles, cells from different embryos. Bars, mean ± s.e.m. P values (unpaired, two-sided *t*-test) shown if $P < 0.05$.

*Ripk1*$^{D325A/+}$ mice were more sensitive than wild-type mice to TNF toxicity (Fig. 4a). All *Ripk1*$^{D325A/+}$ mice ($n = 8$) succumbed within 4 h of TNF treatment, whereas only 50% of *Ripk1*$^{+/+}$ littermates (5 out of 10) succumbed, and at later times, between 5 and 8 h after treatment. After 2 h, *Ripk1*$^{D325A/+}$ mice had increased serum chemokines (Fig. 4b) and increased RIPK1 autophosphorylation in their ileum (Fig. 4c). Increased TNF-induced apoptosis rather than necroptosis caused lethality in the *Ripk1*$^{D325A/+}$ mice, because *Ripk1*$^{D325A/+}$ *Mlkl*$^{-/-}$ mice succumbed as quickly as *Ripk1*$^{D325A/+}$ mice after TNF dosing, whereas *Ripk1*$^{D325A/+}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice all survived (Fig. 4d). The kinase activity of RIPK1 is required for TNF-induced lethality in wild-type mice[19], and this was also true in the *Ripk1*$^{D325A/+}$ mice because *Ripk1*$^{D138N,\,D325A/D138N}$ mice expressing catalytically inactive RIPK1 with one allele also bearing the D325A mutation were completely resistant to TNF-induced lethality (Fig. 4e).

Collectively, our data indicate that Asp325 in RIPK1 is crucial for preventing aberrant cell death during development and for preventing TNF-induced cell death in various cell types. Given that caspase-8 cleaves RIPK1 at this site[14], and that caspase-8 deficiency causes

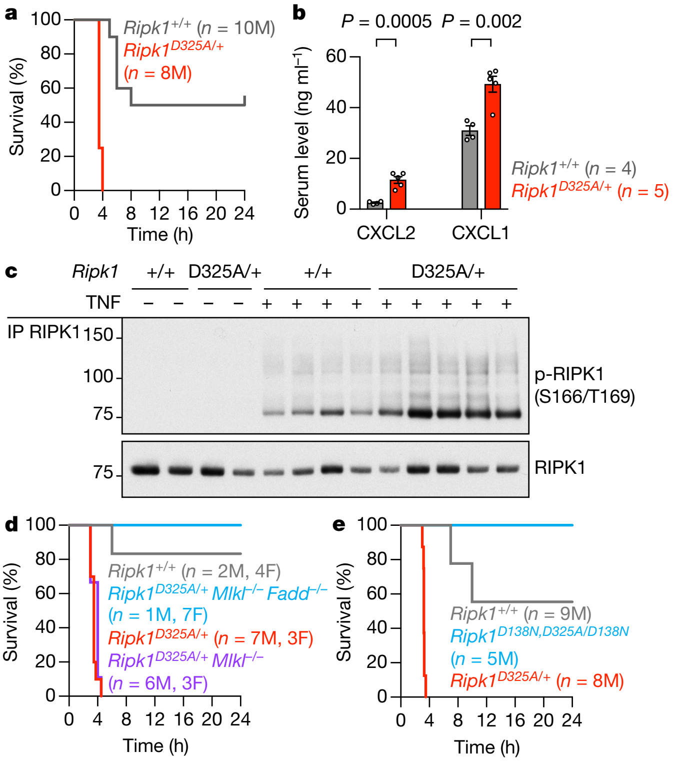


**Fig. 4 | *Ripk1*$^{D325A/+}$ mice require RIPK1 enzymatic activity and FADD for TNF sensitivity. a**, **d**, **e**, Kaplan–Meier survival curves of 6- to 14-week-old mice after dosing with TNF. M, male. F, female. By two-sided log-rank test, *Ripk1*$^{D325A/+}$ differed from *Ripk1*$^{+/+}$ ($P = 0.00001$ (**a**), $P = 0.0001$ (**d**), $P = 0.00002$ (**e**)), *Ripk1*$^{D325A/+}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ ($P = 0.00002$ (**d**)) and *Ripk1*$^{D138N,D325A/D138N}$ ($P = 0.0008$ (**e**)), but not from *Ripk1*$^{D325A/+}$ *Mlkl*$^{-/-}$ ($P = 0.15$ (**d**)). **b**, Serum chemokines in 8- to 9-week-old male littermates, 2 h after TNF treatment. Circles, individual mice. Bars, mean ± s.e.m. Unpaired, two-tailed *t*-test. **c**, Western blots of ilea from 8- to 9-week-old male littermates (*Ripk1*$^{+/+}$, $n = 2$ (untreated), $n = 4$ (2 h TNF); *Ripk1*$^{D325A/+}$, $n = 2$ (untreated), $n = 5$ (2 h TNF)). For gel source data, see Supplementary Fig. 1.

RIPK1-dependent embryonic lethality in mice[21,22], it is likely that RIPK1 Asp325 is cleaved by caspase-8 in cell-death signalling complexes and that cleavage of RIPK1 causes these complexes to fall apart, thereby terminating the death signal. We cannot, however, formally exclude the possibility that RIPK1(D325A) simply adopts a conformation that better interacts with FADD and caspase-8 to promote cell death. Regardless, the identification of Asp325 as a key residue in RIPK1 for preventing inappropriate TNF-induced cell death may explain the periodic fever syndrome observed in humans with heterozygous mutations that affect the equivalent residue in human RIPK1 (N. Lalaoui et al., manuscript submitted).

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1548-0.

Received: 11 February 2019; Accepted: 7 August 2019;

1. Kaiser, W. J. et al. RIP3 mediates the embryonic lethality of caspase-8-deficient mice. *Nature* **471**, 368–372 (2011).
2. Oberst, A. et al. Catalytic activity of the caspase-8-FLIP(L) complex inhibits RIPK3-dependent necrosis. *Nature* **471**, 363–367 (2011).
3. Alvarez-Diaz, S. et al. The pseudokinase MLKL and the kinase RIPK3 have distinct roles in autoimmune disease caused by loss of death-receptor-induced apoptosis. *Immunity* **45**, 513–526 (2016).
4. Chun, H. J. et al. Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature* **419**, 395–399 (2002).
5. Lehle, A. S. et al. Intestinal inflammation and dysregulated immunity in patients with inherited caspase-8 deficiency. *Gastroenterology* **156**, 275–278 (2019).
6. Varfolomeev, E. E. et al. Targeted disruption of the mouse caspase 8 gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* **9**, 267–276 (1998).
7. Ch'en, I. L., Tsau, J. S., Molkentin, J. D., Komatsu, M. & Hedrick, S. M. Mechanisms of necroptosis in T cells. *J. Exp. Med.* **208**, 633–641 (2011).
8. Kaiser, W. J. et al. Toll-like receptor 3-mediated necrosis via TRIF, RIP3, and MLKL. *J. Biol. Chem.* **288**, 31268–31279 (2013).
9. Kang, T. B. et al. Caspase-8 serves both apoptotic and nonapoptotic roles. *J. Immunol.* **173**, 2976–2984 (2004).
10. Webster, J. D., Solon, M., Haller, S. & Newton, K. Detection of necroptosis by phospho-RIPK3 immunohistochemical labeling. *Methods Mol. Biol.* **1857**, 153–160 (2018).
11. Medema, J. P. et al. FLICE is activated by association with the CD95 death-inducing signaling complex (DISC). *EMBO J.* **16**, 2794–2804 (1997).
12. Pop, C. et al. FLIP$_L$ induces caspase 8 activity in the absence of interdomain caspase 8 cleavage and alters substrate specificity. *Biochem. J.* **433**, 447–457 (2011).
13. O'Donnell, M. A. et al. Caspase 8 inhibits programmed necrosis by processing CYLD. *Nat. Cell Biol.* **13**, 1437–1442 (2011).
14. Lin, Y., Devin, A., Rodriguez, Y. & Liu, Z. G. Cleavage of the death domain kinase RIP by caspase-8 prompts TNF-induced apoptosis. *Genes Dev.* **13**, 2514–2526 (1999).
15. Feng, S. et al. Cleavage of RIP3 inactivates its caspase-independent apoptosis pathway by removal of kinase domain. *Cell. Signal.* **19**, 2056–2067 (2007).
16. Kang, T. B. et al. Mutation of a self-processing site in caspase-8 compromises its apoptotic but not its nonapoptotic functions in bacterial artificial chromosome-transgenic mice. *J. Immunol.* **181**, 2522–2532 (2008).
17. Watanabe-Fukunaga, R., Brannan, C. I., Copeland, N. G., Jenkins, N. A. & Nagata, S. Lymphoproliferation disorder in mice explained by defects in Fas antigen that mediates apoptosis. *Nature* **356**, 314–317 (1992).
18. Zhang, X., Dowling, J. P. & Zhang, J. RIPK1 can mediate apoptosis in addition to necroptosis during embryonic development. *Cell Death Dis.* **10**, 245 (2019).
19. Newton, K. et al. Activity of protein kinase RIPK3 determines whether cells die by necroptosis or apoptosis. *Science* **343**, 1357–1360 (2014).
20. Kaiser, W. J. et al. RIP1 suppresses innate immune necrotic as well as apoptotic cell death during mammalian parturition. *Proc. Natl Acad. Sci. USA* **111**, 7753–7758 (2014).
21. Rickard, J. A. et al. RIPK1 regulates RIPK3–MLKL-driven systemic inflammation and emergency hematopoiesis. *Cell* **157**, 1175–1188 (2014).
22. Dillon, C. P. et al. RIPK1 blocks early postnatal lethality mediated by caspase-8 and RIPK3. *Cell* **157**, 1189–1202 (2014).
23. Newton, K. et al. RIPK1 inhibits ZBP1-driven necroptosis during development. *Nature* **540**, 129–133 (2016).
24. Gong, Y. N. et al. ESCRT-III acts downstream of MLKL to regulate necroptotic cell death and its consequences. *Cell* **169**, 286–300.e16 (2017).
25. Mandal, P. et al. RIP3 induces apoptosis independent of pronecrotic kinase activity. *Mol. Cell* **56**, 481–495 (2014).
26. Dondelinger, Y. et al. RIPK3 contributes to TNFR1-mediated RIPK1 kinase-dependent apoptosis in conditions of cIAP1/2 depletion or TAK1 kinase inhibition. *Cell Death Differ.* **20**, 1381–1392 (2013).
27. Holler, N. et al. Fas triggers an alternative, caspase-8-independent cell death pathway using the kinase RIP as effector molecule. *Nat. Immunol.* **1**, 489–495 (2000).

## METHODS

**Mice.** All mouse studies complied with relevant ethical regulations and were approved by the Genentech institutional animal care and use committee. *Ripk1*$^{D138N/D138N}$, *Ripk3*$^{-/-}$, and *Casp8*$^{+/-}$ mice[19], *Mlkl*$^{-/-}$ mice[28], and *Tnfr1*$^{-/-}$ mice[29] have previously been described. *Tnfr1*$^{-/-}$ mice were backcrossed to a C57BL/6N background for at least 13 generations. *Casp8*$^{C362A/+}$ (Extended Data Fig. 1a) and *Fadd*$^{+/-}$ (Extended Data Fig. 3d) mice were generated by Taconic (Germany) using C57BL/6 NTac embryonic stem (ES) cells. The *Fadd* deleted region corresponds to genomic position chr 7:144,581,754–144,584,070 (GRCm38.p4 C57BL/6J). *Casp8* codon TGC (encoding Cys362) was mutated to GCA (Ala). *Cyld*$^{D215A/+}$ mice (Extended Data Fig. 2e) were generated by Ozgene (Australia) from C57BL/6 ES cells. *Cyld* codon GAT (Asp215) was mutated to GCT (Ala).

*Casp8*$^{1\times DA/+}$ mice were generated at Genentech using C57BL/6 C2 ES cells and had *Casp8* codon GAT (Asp387) mutated to GCC (Ala). *Casp8*$^{1\times DA/1\times DA}$ zygotes were then modified using *Cas9* mRNA (ThermoFisher A29378), a sgRNA (Synthego) targeting the sequence 5′ GTA AAC TTT GTC TGA AGT C with protospacer adjacent motif (PAM) TGG, and oligonucleotide template (5′CTG AGC ACT TGG ACA TAG CCA GTG TCT GAG CTG CAG CAT CAA CCC CTG GAT TGG GCT TGT GTT TTC CAG ACC TCA GCT AAA GTT TAC CAA ATG AAG AAC AAA CCT CGG GGA TAC TGT CTG ATC ATC AAC AAT CAT GAT TTC AGC AAG GCC CGG; Integrated DNA Technologies), which were injected into the pronucleus to mutate codons ACT TCA GAC (Thr-Ser-Asp225) to ACC TCA GCT (Thr-Ser-Ala225) (sequence underlined in the oligonucleotide template). Mosaic founders containing the desired mutations were screened using next-generation sequencing at the top 15 off-target loci predicted by the CRISPR gRNA design tool Benchling. Founders lacking alterations at these off-target loci were then backcrossed to *Casp8*$^{1\times DA/1\times DA}$ mice for germline transmission of the edited chromosome. Another round of gene editing was then performed with Cas9, a sgRNA targeting 5′CAA GCT AGT GAG TCA CGG GT with PAM AGG, and oligonucleotide template (5′AGT GTG ACG TTT TTT TGG TTG CTT GCA GAG ATG AGC CTC AAA ATG GCG GAA CTG TGT GAC TCG CCA AGA GAA CAA GCT AGT GAG TCA CGC GTA GGT GTG TCT CCT ACC TCT CTC TTT GCA TTG GTG TTC CTG TTT CCT TTG GTT GGT TCC TTT) to mutate codons GAC AGT GAG TCA CGG (Asp218-Ser-Glu-Ser-Arg) to GCT AGT GAG TCA CGC (Ala218-Ser-Glu-Ser-Arg). The *Casp8*$^{D218A, D225A, D387A}$ allele was abbreviated to *Casp8*$^{3\times DA}$. *Casp8*$^{3\times DA/3\times DA}$ zygotes were then edited with Cas9, an sgRNA targeting 5′ AGA GAT GAG CCT CAA AAT GG and PAM CGG, and oligonucleotide template (5′AAA CAG GAA CAC CAA TGC AAA GAG AGA GGT AGG AGA CAC ACC TAC CCG TGA CTC ACT AGC TTG TTC TCT TGG CGA GGC ACA GAG CTC CGC CAT CTT CAG GCT CAT CTC TGC AAG CAA CCA AAA AAA CGT CAC ACT TAG GAC AAC CTA AGG ACC) to mutate codons CTC AAA ATG GCG GAA CTG TGT GAC TCG CCA AGA GAA CAA GCT AGT GAG TCA CGC (Leu-Lys-Met-Ala-Glu-Leu-Cys-Asp212-Ser-Pro-Arg-Glu-Gln-Ala218-Ser-Glu-Ser-Lys) to CTG AAG ATG GCG GAG CTC TGT GCC TCG CCA AGA GAA CAA GCT AGT GAG TCA CGG (Leu-Lys-Met-Ala-Glu-Leu-Cys-Ala212-Ser-Pro-Arg-Glu-Gln-Ala218-Ser-Glu-Ser-Lys). This *Casp8*$^{D212A, D218A, D225A, D387A}$ allele was abbreviated to *Casp8*$^{4\times DA}$.

*Ripk1*$^{D325A/+}$ mice and *Ripk1*$^{D138N,D325A/D138N}$ mice were generated at Genentech using C57BL/6N and *Ripk1*$^{D138N/D138N}$ zygotes, respectively. *Cas9* mRNA, an sgRNA targeting 5′TTT GAC CTG CTC GGA GGT AA with PAM GGG, and oligonucleotide template (5′ GTA CAT GTC TTC TCA TTA CAG AAA GAG TAT CCA GAT CAA AGC CCA GTG CTG CAG AGA ATG TTT TCA CTG CAG CAT GCC TGC GTT CCC TTA CCA CCA AGC AGG TCA AAT TCA GGT AAC TCA CCT ATT CGT TCA TTT GCA TAC TCG CTC ACT TAC) were used to mutate *Ripk1* codons GAC TGT GTA CCC TTA CCT CCG (Asp325-Cys-Val-Pro-Leu-Pro-Pro) to GCC TGC GTT CCC TTA CCA CCA (Ala325-Cys-Val-Pro-Leu-Pro-Pro).

*Cflar*$^{D371A, D377A/+}$ mice were generated at Genentech. C57BL/6N zygotes were injected with *Cas9* mRNA, an sgRNA targeting 5′ TGA TGG CCC ATC TAC CTC C, and PAM AGG, and an oligonucleotide template (5′ ACG TGC CCT TCT CTC AGA GGG AAG CCA AAG CTC TTT TTT ATT CAG AAC TAT GAG TCG TTA GGT AGC CAG TTG GAA GCA AGT AGT CTA GAA GTT GCA GGG CCA TCA ATA AAA AAT GTG GAC TCT AAG CCC CTG CAA CCC AGA CAC TGC ACA ACT CAC CCA GAA GCT GAT ATC) to mutate codons GAT AGC AGC CTG GAG GTA GAT (Asp371-Ser-Ser-Leu-Glu-Val-Asp377) to GCA AGT AGT CTA GAA GTT GCA (Ala371-Ser-Ser-Leu-Glu-Val-Ala377). This *Cflar*$^{D371A, D377A}$ allele was abbreviated to *Cflar*$^{2\times DA}$.

All alleles were maintained on a C57BL/6N genetic background. For timed pregnancies, mice were designated E0.5 on the morning a vaginal plug was detected. Mice were designated P0 on the day their birth was detected. Where indicated, mice were injected intravenously with 500 μg mouse TNF (Genentech) per kg body weight. Calculations to determine group sizes were not performed, mice were not randomized but were grouped according to genotype, and experiments were unblinded. Body temperatures were measured using a lubricated digital rectal probe. Mice were euthanized if severely lethargic or if their body temperature dropped to 25 °C. Serum samples were analysed by Luminex assay using a mouse cytokine/chemokine 32-plex panel (EMD Millipore).

**Genotyping.** Primers for *Casp8*$^{C362A/+}$ mice (5′GAC AAA GAC AAG GAA ATG GAC C and 5′ACT TTG CTG AGT TCT TGC ACC) amplified 282-bp wild-type and 351-bp knock-in DNA fragments; primers for *Fadd*$^{+/-}$ mice (5′CTG AGC TAC ATG GTG TGA AG, 5′ACT GGA CTG GCT TAC TTG A, and 5′AAC AGA AGT AAC TGC CGA TAA C) amplified 238-bp wild-type and 326-bp knockout DNA fragments; primers for *Casp8*$^{1\times DA/+}$ mice (5′CCT ACC GAG ATC CTG TGA ATG G, 5′AGA GAA TAG GAA CTT CCA TGC CAA and 5′CAC TGA CCT GGG AAA TAC AAC TG) amplified 255-bp wild-type, and 216-bp and 289-bp knock-in DNA fragments; primers for *Casp8*$^{3\times DA/+}$ mice (5′AAG CTT CCG AGT CAC G, 5′AAG AGA ACA AGA CAG TGA GT, and 5′CTC CAG AGT TTC CAA AAG G) amplified 139-bp wild-type and 131-bp knock-in DNA fragments; primers for *Casp8*$^{4\times DA/+}$ mice (5′GTG TGA CTC GCC AAG A, 5′AGC TCT GTG CCT CGC, and 5′TTT CTC TCC AGA GTT TCC AAA) amplified 156-bp wild-type and 160-bp knock-in DNA fragments; primers for *Cflar*$^{2\times DA/+}$ mice (5′TCA AGA GAC CTT CAC TTC C, 5′CCT CCA GGC TGC TAT C, and 5′GAC TAC TTG CTT CCA ACT G) amplified 320-bp wild-type and 314-bp knock-in DNA fragments; primers for *Cyld*$^{D215A/+}$ mice (5′ACG ATG AGG TAA TTT CAC TGT, 5′AGT TGC CAA TAG GGT TAT CC, and 5′ATT GGC AAT GAA ACT AAT ACA TAT) amplified 536-bp wild-type and 589-bp knock-in DNA fragments.

The *Ripk1*$^{D325A}$ allele was genotyped by digital-droplet (dd)PCR. Genomic DNA was prepared using the DNeasy Blood & Tissue kit (Qiagen 69506) and diluted to 30 ng/μl. Primers (*Ripk1*-F, 5′ACT GGT TAC TAA AGT GTG TAC; *Ripk1*-R, 5′GAA CGA ATA GGT GAG TTA CC; *Apob*-F, 5′CAC GTG GGC TCC AGC ATT; *Apob*-R, 5′-TCA CCA GTC ATT TCT GCC TTT G; 25 μM stock concentration) and probes (*Ripk1* knock-in 5′-[6-FAM]-CAG CAT GCC TGC GTT CCC-3′-[Iowa Black FQ] and *Apob* 5′-[HEX]-CC+ A+A TGG +TC-[ZEN]-GGG CAC +TG-3′-[Iowa Black FQ], or *Apob* 5′-[6-FAM]-CC+A+ATGG+TC-[ZEN]-GGG CAC+TG-3′-[Iowa Black FQ] and *Ripk1* WT 5′-[HEX]-CAG CAT +GA +CT +GT GTA CCC-3′-[Iowa Black FQ]; stock concentration 6.25 μM) were from Integrated DNA Technologies. + indicates a locked nucleic acid base. *Apob* served as a reference gene. Reactions (2 per sample) contained 12.5 μl ddPCR Supermix for Probes (no dUTP) (Bio-Rad 186-3023), 1 μl of each of the 4 primers, 1 μl of each of the 2 probes and 1 μl DNA in a final volume of 25 μl. Droplets were generated on the QX200 AutoDG Droplet Digital PCR System (Bio-Rad Laboratories). Plates were heat-sealed with foil and placed in a C1000 Touch Thermal Cycler (Bio-Rad Laboratories) with the following cycling conditions: 95 °C for 10 min; 40× (94 °C for 30 s, 58 °C for 30 s; 98 °C for 10 min) then kept at 4 °C. Droplets were scanned on the QX200 Droplet Reader (Bio-Rad Laboratories) and copy numbers were assessed with QuantaSoft software. At least 10,000 droplets were counted per sample.

**Cells.** Thymocytes, bone marrow cells and primary MEFs were cultured in the high-glucose version of Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum, 2 mM glutamine, 10 mM HEPES (pH 7.2), 1 × non-essential amino acids solution, 100 U/ml penicillin and 100 μg/ml streptomycin. MEFs in Fig. 2 and Extended Data Fig. 5a, b were isolated from E10.5 embryos and grown on tissue culture dishes pre-coated with 0.1% gelatin in PBS. All other MEFs were from E13.5 or E14.5 embryos. Primary macrophages were differentiated from bone marrow cells in non-treated plates in medium supplemented with 25 ng/ml M-CSF (R&D Systems) for 5–6 d.

Mouse cells were stimulated with 100 ng/ml mouse TNF (Genentech), 20 μM Z-VAD-FMK (Promega), 100 ng/ml FasL (Enzo Life Sciences), 100 ng/ml TRAIL (Enzo Life Sciences), 250 ng/ml (Fig. 3a, b), 1 μg/ml (Extended Data Fig. 1b) or 10 μg/ml (Extended Data Fig. 1c) cycloheximide (Sigma), 30 μM Nec-1 (Tocris Bioscience), 100 nM 5Z-7-oxozeaenol (Sigma) and 1 μM GSK′843[25]. To assess cell viability, adherent MEFs and bone-marrow-derived macrophages (BMDMs) were detached with trypsin, pooled with floating dead cells, stained with 2.5 μg/ml propidium iodide (PI; BD Biosciences), and analysed in a FACSCanto II cytometer (BD Biosciences) with BD FACSDiva 8.0. Viable cells were PI-negative. Thymocytes in suspension were stained directly with PI. Caspase-8 activity was measured using the Caspase-Glo 8 assay (Promega).

Mesenteric lymph node cells were stained with APC-Cy7- or FITC-conjugated 145-2C11 anti-CD3ε (BD Biosciences) and FITC- or APC-conjugated RA3-6B2 anti-B220 (BD Biosciences) antibodies in the presence of 2% normal rat serum and 1 μg/ml 2.4G2 anti-CD16/CD32 (BD Biosciences). Leukocytes were identified based on their forward scatter (FSC) and side scatter profiles. Dead cells that stained with 7-AAD (BD Biosciences), plus doublets, identified by their FSC-A versus FSC-W profiles, were excluded from analyses as indicated in Extended Data Fig. 1l. Data were acquired using a BD FACSCantoII cytometer (BD Biosciences) and analysed with FlowJo 10.3.

293T cells (ATCC CRL-3216; tested for mycoplasma but not authenticated) were transfected with FuGENE HD transfection reagent (Promega), mouse caspase-8 cloned into pCMV-3tag7 (Agilent Technologies) with a N-terminal 3× Myc tag, and mouse RIPK1 (wild-type or D325A) cloned into pCMV-3tag6 (Agilent Technologies) with a N-terminal 3× Flag tag for 18 h.

**Immunoprecipitation and western blotting.** Cells were lysed in 20 mM Tris.HCl pH 7.5, 135 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 1% Triton X-100, 10% glycerol, phosSTOP phosphatase inhibitor (Roche), and Halt protease inhibitor cocktail (Pierce). Insoluble material was removed by centrifugation at 20,000*g* before addition of LDS sample buffer or immunoprecipitation. In the experiments shown in Fig. 3a, Extended Data Fig. 5f, cells were lysed in 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 1% NP40, phosSTOP phosphatase inhibitor (Roche), Halt protease inhibitor cocktail (Pierce), and DNase (about 80 U/ml, Qiagen). The whole-cell lysate was denatured directly in LDS sample buffer.

RIPK1 was immunoprecipitated with rat 10C7 anti-mouse RIPK1 antibody (Genentech) conjugated to Dynabeads M-270 epoxy beads (Invitrogen). Western blot antibodies recognized RIPK3 (1G6, Genentech), phosphorylated RIPK3(Thr231/Ser232) (GEN135-35-9, Genentech), RIPK1 (610459, BD Biosciences, or 10C7, Genentech), phosphorylated RIPK1(Ser166/Thr169) (GEN150-33-4, Genentech), β-actin (MP Biomedicals, mouse clone C4), FADD (1.28E12, Genentech), caspase-8 (1G12, Enzo Life Sciences), c-FLIP (2.21H2, Genentech), CYLD (GN55-9, Genentech), MLKL (3H1, Millipore), phosphorylated MLKL Ser345 (ab196436, Abcam), Flag (A8592, Sigma), and Myc (GTX21261, Genetex). Other antibodies were from Cell Signaling Technologies: p-ERK (9101), ERK (9102), p-JNK (4668), JNK (9258), p-IκBα (2859), IκBα (9242), p-p65/RelA (3033), p65/RelA (8242), p-p38 (9211), p38 (8690), caspase-3 (9662), cleaved caspase-3 (9664) and cleaved caspase-8 (8592).

**Immunohistochemistry and immunofluorescence.** Paraformaldehyde-fixed yolk sacs and formalin-fixed, paraffin-embedded tissue sections were labelled with antibodies detecting cleaved caspase-3[19], PECAM-1[19], or p-RIPK3(T231/S232)[10] as described. Caspase-8 was labelled with 2.5 μg/ml 1G12 rat anti-caspase 8 antibody (Enzo) on a Dako autostainer with pH 6 target retrieval (Dako), hydrogen peroxidase, and ScyTek biotin block. Bound antibody was detected with rabbit anti-rat secondary and the ABC-Peroxidase Elite KIT (Vector Labs) with DAB chromogen. RIPK1 was labelled with 1.5 μg/ml 10C7 rat anti-RIPK1 antibody (Genentech) on a Dako autostainer with pH 6 target retrieval, hydrogen peroxidase and TNB block, and Powervision polymer-HRP (Leica) detection with DAB.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding authors on reasonable request.

28. Murphy, J. M. et al. The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity* **39**, 443–453 (2013).
29. Peschon, J. J. et al. TNF receptor-deficient mice reveal divergent roles for p55 and p75 in several models of inflammation. *J. Immunol.* **160**, 943–952 (1998).

**Acknowledgements** We thank M. Long, M. Dempsey, A. Verducci, T. Scholl, T. Alcantar, and E. Lu for mouse husbandry, A. Strasser, D. Siler, and the Genentech genetic analysis and histology laboratories for technical assistance, G. Salvesen and D. Vucic for discussions, and W. Alexander (WEHI, Australia) for *Mlkl*$^{-/-}$ mice.

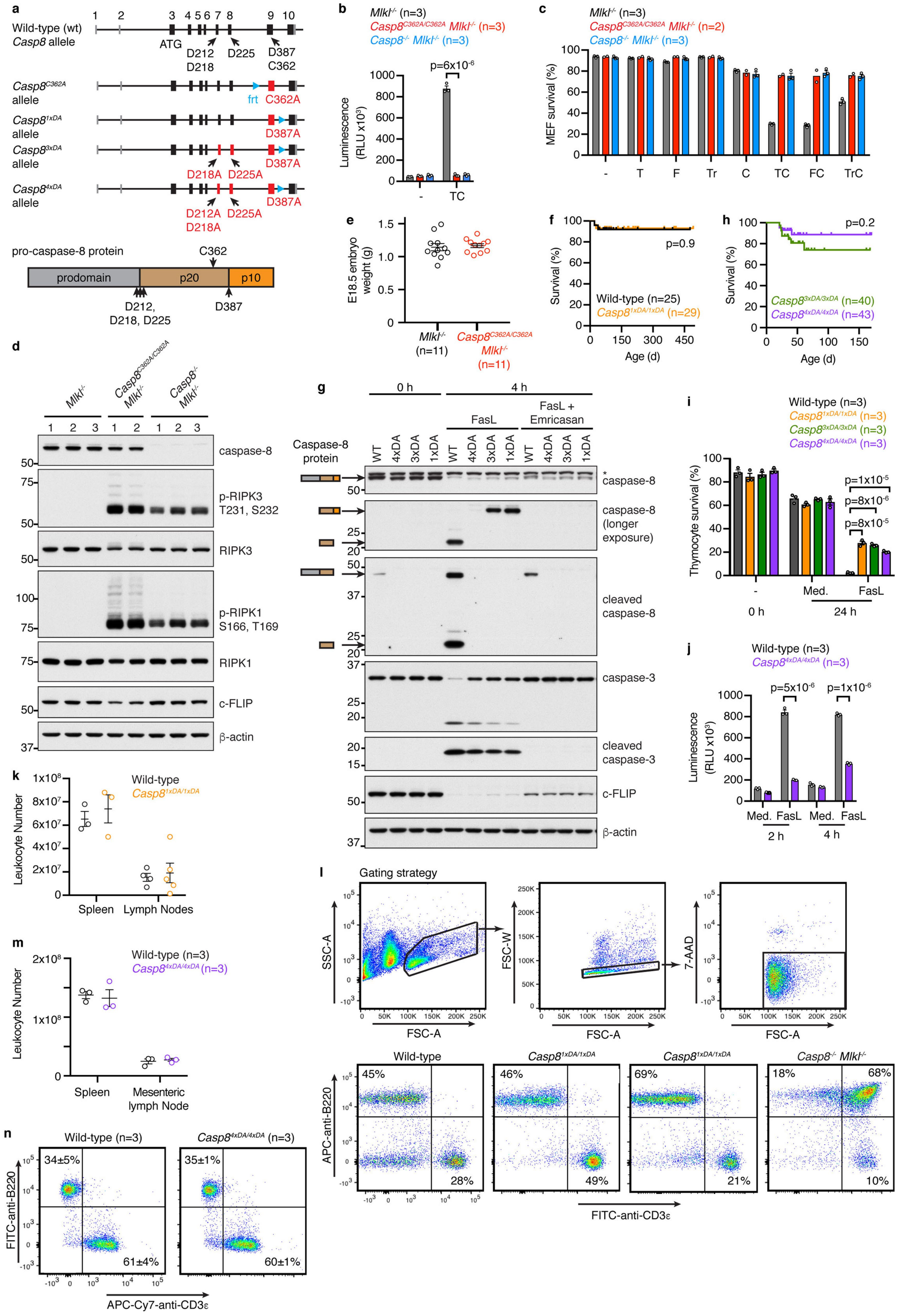
**Author contributions** M.R.-G. generated the *Casp8*$^{1\times DA/+}$, *Casp8*$^{4\times DA/+}$, *Cflar*$^{2\times DA/+}$, and *Ripk1*$^{D325A/+}$ mice; K.N., A.M., D.L.D., and K.E.W. designed and performed experiments; L.K. performed confocal microscopy; M.D. performed immunohistochemistry; J.D.W. analysed histological data; V.M.D. contributed to experimental design; K.N. wrote the paper with input from all authors.

**Competing interests** All authors were employees of Genentech.

**Additional information**

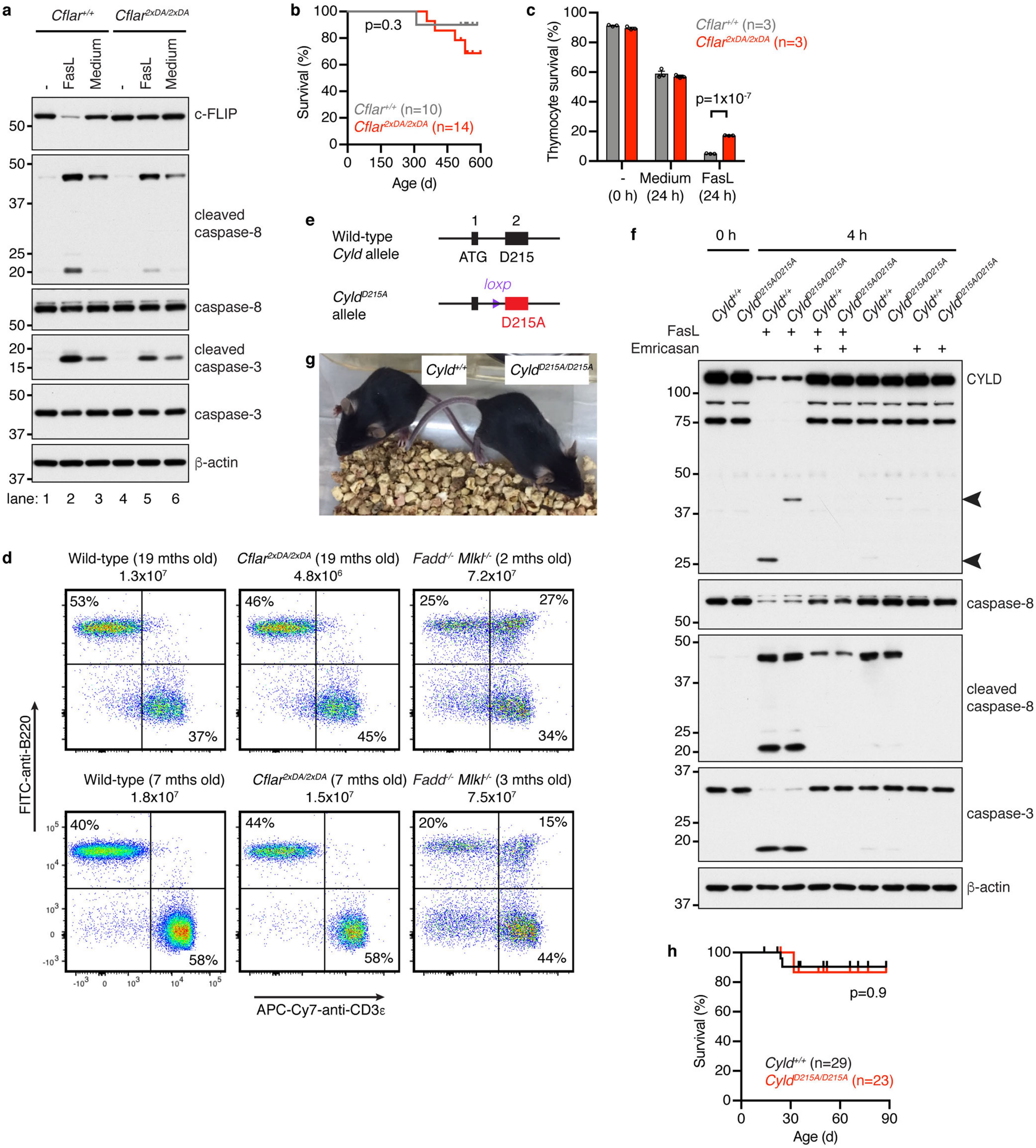
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1548-x.

**Correspondence and requests for materials** should be addressed to K.N. or V.M.D.

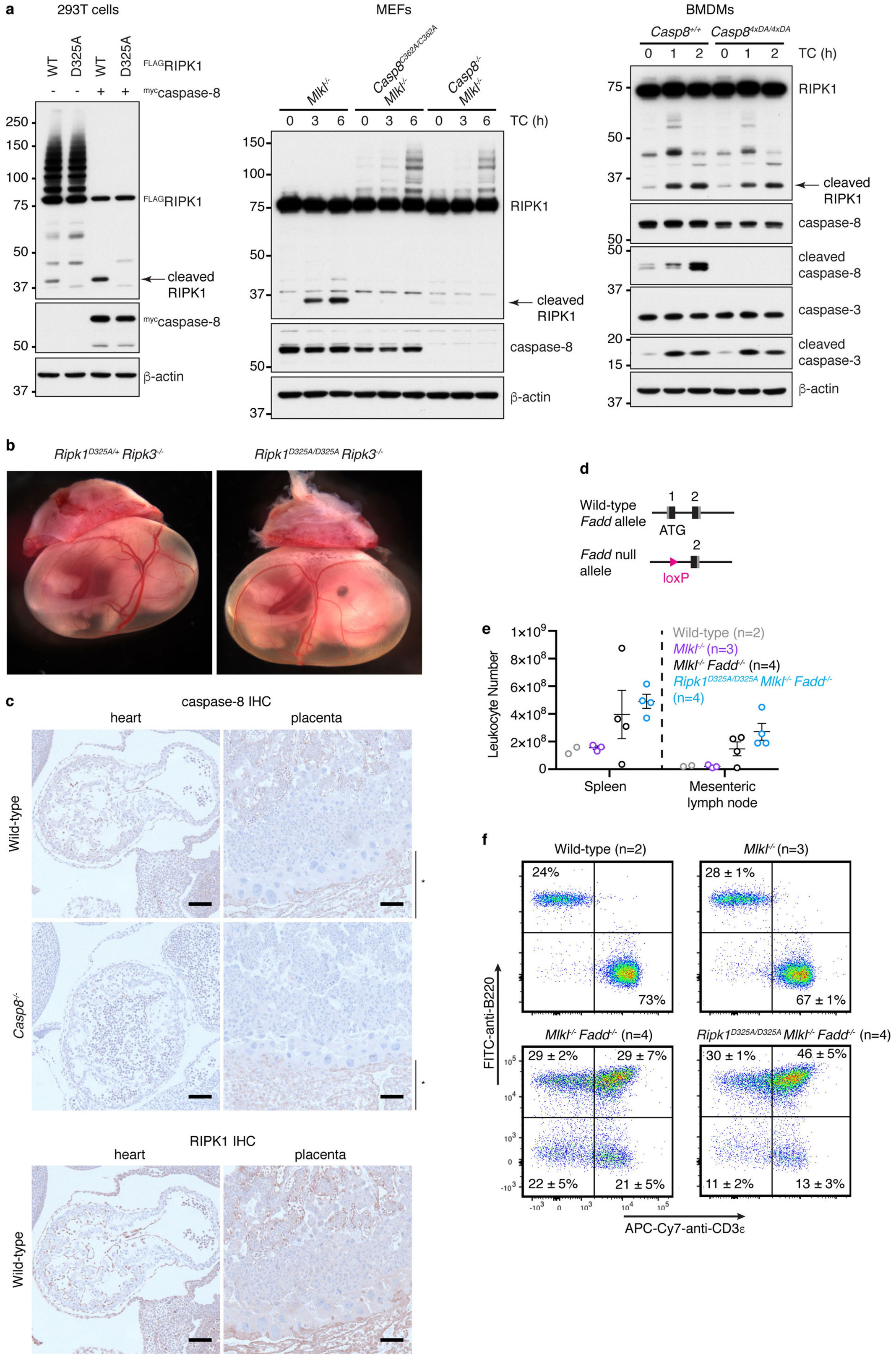


**Extended Data Fig. 1 |** See next page for caption.

**Extended Data Fig. 1 | Catalytic activity of caspase-8 is required to prevent necroptosis, but autoprocessing of caspase-8 is not required.** **a**, Schematic showing the organization of the different *Casp8* alleles and of pro-caspase-8 protein. Non-coding and coding *Casp8* exons are represented by grey and black boxes, respectively. **b**, Graph indicates cleavage of a proluminogenic caspase-8 substrate in primary MEF lysates at 6 h after treatment (−, no treatment; C, cycloheximide; T, TNF). RLU, relative luminescence unit. Circles, cells from different embryos (*Mlkl*$^{-/-}$, $n = 3$; *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$, $n = 3$; *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$, $n = 3$). Bars, mean ± s.e.m. Unpaired two-tailed *t*-test. **c**, Graph indicates the percentage of primary MEFs that were viable 16 h after treatment (F, FasL; Tr, TRAIL). Circles, cells from different embryos (*Mlkl*$^{-/-}$, $n = 3$; *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$, $n = 2$; *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$, $n = 3$). Bars, mean ± s.e.m. **d**, Western blots of primary MEFs. Each lane represents cells from a different embryo (*Mlkl*$^{-/-}$, $n = 3$; *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$, $n = 2$; *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$, $n = 3$). β-Actin loading control performed after caspase-8. **e**, Body weights of E18.5 embryos. Circles, different embryos (*Mlkl*$^{-/-}$, $n = 11$; *Casp8*$^{C362A/C362A}$ *Mlkl*$^{-/-}$, $n = 11$). Lines, mean ± s.e.m. $P = 0.7$; unpaired two-tailed *t*-test. **f**, Kaplan–Meier survival curve of wild-type ($n = 25$) and *Casp8*$^{1\times DA/1\times DA}$ ($n = 29$) littermates. $P = 0.9$; two-sided log-rank test. **g**, Western blots of thymocytes. The pan-caspase inhibitor emricasan revealed caspase-dependent cleavage events. Asterisk, non-specific band detected by the caspase-8 antibody. β-Actin loading control performed after cleaved caspase-8. Results representative of two independent experiments. **h**, Kaplan–Meier survival curve of *Casp8*$^{3\times DA/3\times DA}$ ($n = 40$) and *Casp8*$^{4\times DA/4\times DA}$ ($n = 43$) littermates. $P = 0.2$; two-sided log-rank test. **i**, Graph indicates the percentage of viable thymocytes before treatment (−), and after culture in either medium alone (Med.) or FasL for 24 h. Circles, cells from different mice (wild-type, $n = 3$; *Casp8*$^{1\times DA/1\times DA}$, $n = 3$; *Casp8*$^{3\times DA/3\times DA}$, $n = 3$; *Casp8*$^{4\times DA/4\times DA}$, $n = 3$). Bars, mean ± s.e.m. Unpaired two-tailed *t*-test. **j**, Graph indicates cleavage of a proluminogenic caspase-8 substrate in thymocyte lysates. Circles, cells from different mice (wild-type, $n = 3$; *Casp8*$^{4\times DA/4\times DA}$, $n = 3$). Bars, mean ± s.e.m. Unpaired two-tailed *t*-test. **k**, Graph indicates leukocyte numbers in spleen and lymph nodes (axillary, brachial, inguinal, and mesenteric) of mice aged 12–16 months (spleen, $n = 3$ per genotype; lymph nodes, $n = 4$ wild-type, $n = 5$ *Casp8*$^{1\times DA/1\times DA}$). Circles, cells from different mice. Lines, mean ± s.e.m. **l**, Flow cytometric analysis of lymph node cells from 1 wild-type and 2 *Casp8*$^{1\times DA/1\times DA}$ mice aged 16 months, and as a control, one *Casp8*$^{-/-}$ *Mlkl*$^{-/-}$ mouse aged 3 months (this strain develops lymphadenopathy due to impaired Fas-induced apoptosis[3]). **m**, Graph indicates leukocyte numbers in spleen and mesenteric lymph node of mice aged 4–5 months ($n = 3$ mice per genotype). Circles, cells from different mice. Lines, mean ± s.e.m. **n**, Flow cytometric analysis of mesenteric lymph node cells from the mice in **m**. Percentages represent the mean ± s.e.m. The three gates in **l** were also applied to these samples. For gel source data, see Supplementary Fig. 1.

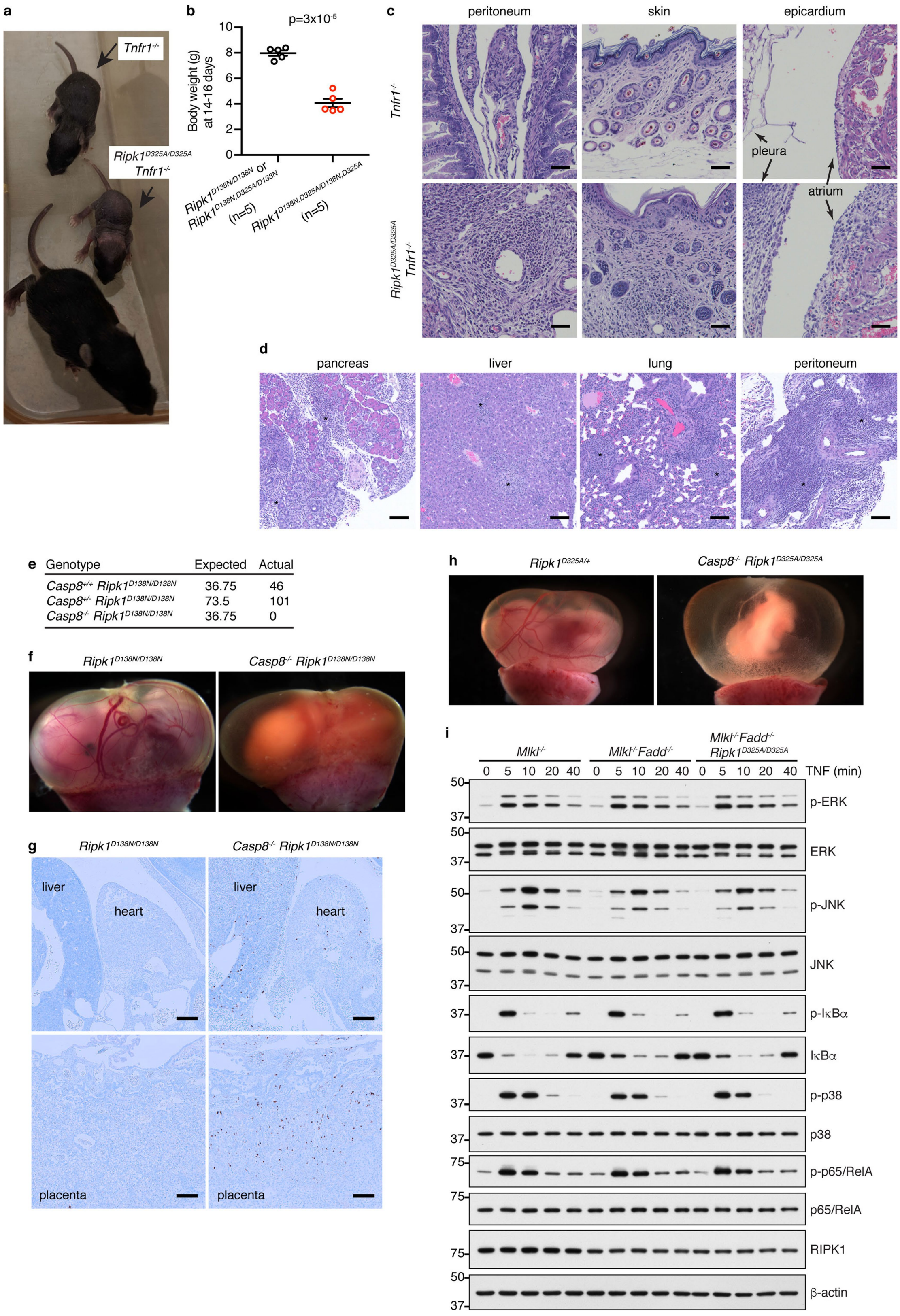


**Extended Data Fig. 2 | Characterization of mice expressing c-FLIP(D371A/D377A) or CYLD(D215A).** **a**, Western blots of thymocytes that were freshly isolated (−), or cultured in either medium alone or FasL for 2.5 h. β-Actin loading control performed after caspase-8. Results representative of two independent experiments. **b**, Kaplan–Meier survival curve of *Cflar*$^{+/+}$ ($n = 10$) and *Cflar*$^{2\times DA/2\times DA}$ littermates ($n = 14$). $P = 0.3$, two-sided log-rank test. **c**, Graph shows the percentage of thymocytes that were viable 24 h after treatment. Circles, cells from different mice (*Cflar*$^{+/+}$, $n = 3$; *Cflar*$^{2\times DA/2\times DA}$, $n = 3$). Bars, mean ± s.e.m. Unpaired two-tailed $t$-test. **d**, Flow cytometric analysis of mesenteric lymph node cells of wild-type ($n = 2$) and *Cflar*$^{2\times DA/2\times DA}$ ($n = 2$) littermates, and as controls, *Fadd*$^{-/-}$ *Mlkl*$^{-/-}$ mice ($n = 2$). Numbers below each genotype indicate total lymph node cellularity. Analyses used the gating strategy shown in Extended Data Fig. 11. **e**, Schematic showing the organization of the *Cyld*$^{D215A}$ allele. **f**, Western blots of thymocytes. Arrowheads highlight caspase-dependent CYLD cleavage products. β-Actin loading control performed after cleaved caspase-8. Results representative of two independent experiments. **g**, Littermate males aged 18 days. Results representative of 29 wild-type and 23 *Cyld*$^{D215A/D215A}$ mice. **h**, Kaplan–Meier survival curve of *Cyld*$^{+/+}$ ($n = 29$) and *Cyld*$^{D215A/D215A}$ ($n = 23$) littermates. $P = 0.9$, two-sided log-rank test. For gel source data, see Supplementary Fig. 1.



**Extended Data Fig. 3** | See next page for caption.

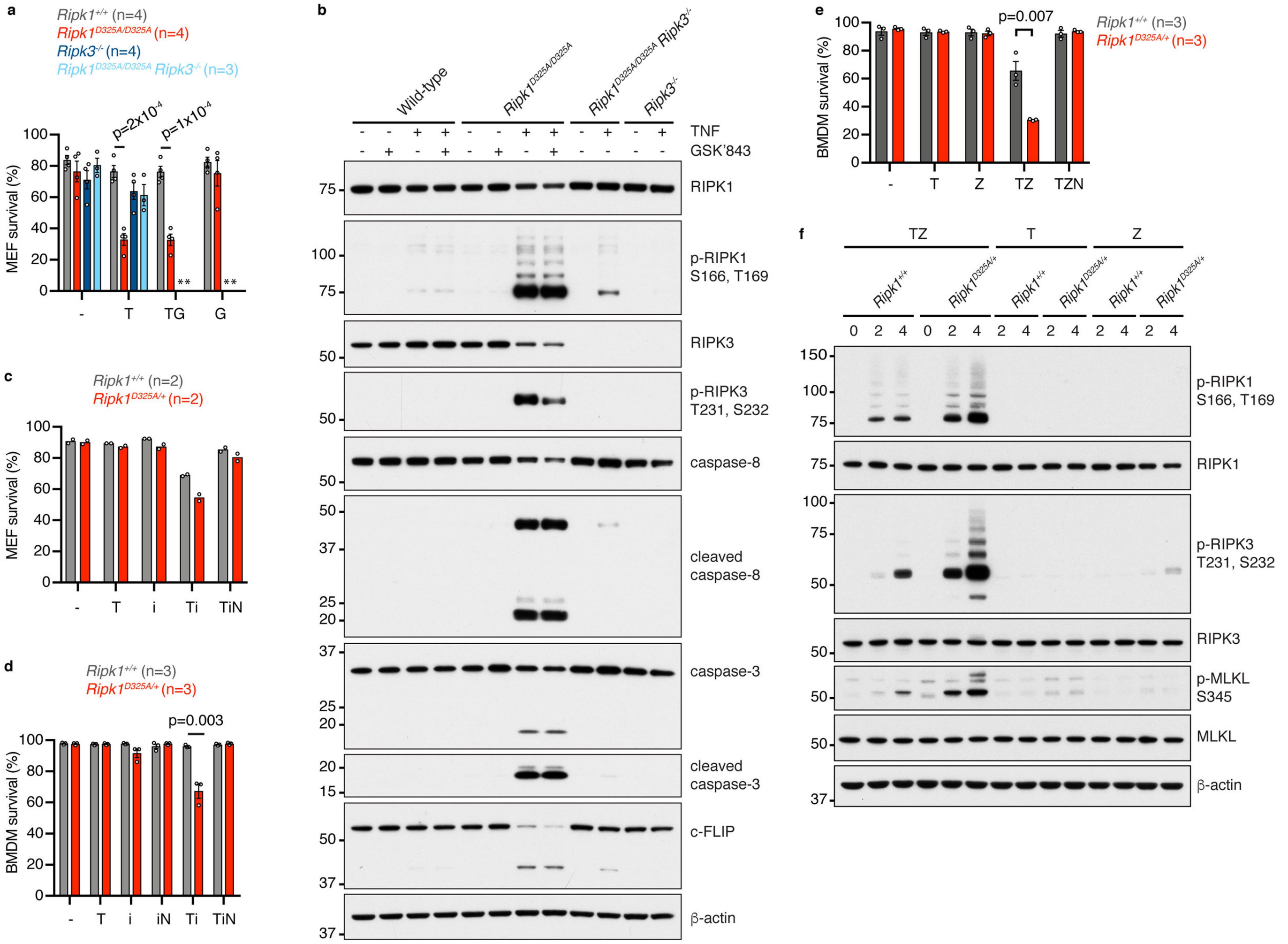
**Extended Data Fig. 3 | *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice develop splenomegaly and lymphadenopathy similar to *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$ mice. a**, Western blots of 293T cells overexpressing mouse caspase-8 and mouse RIPK1 (left), primary MEFs (middle), or BMDMs (right). Results representative of two independent experiments. The N-terminal RIPK1 cleavage product appears larger in the 293T cells owing to the N-terminal Flag tag. T, TNF; C, cycloheximide. β-Actin loading control performed after MYC (left) or caspase-8 (middle, right). For gel source data, see Supplementary Fig. 1. **b**, E12.5 littermates representative of six *Ripk1*$^{D325A/+}$ *Ripk3*$^{-/-}$ and four *Ripk1*$^{D325A/D325A}$ *Ripk3*$^{-/-}$ embryos. **c**, E10.5 embryo sections immunolabelled for either RIPK1 or caspase-8 (brown). Results representative of five wild-type and six *Casp8*$^{-/-}$ embryos. Scale bar, 100 μm. Asterisk shows maternally derived *Casp8*$^{+/-}$ decidua in the placenta. **d**, Schematic showing the organization of the *Fadd* knockout allele. **e**, Leukocyte numbers in mice aged 10–12 weeks (wild-type, $n = 2$; *Mlkl*$^{-/-}$, $n = 3$; *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$, $n = 4$; *Ripk1*$^{D325A/D325A}$ *Mlkl*$^{-/-}$ *Fadd*$^{-/-}$, $n = 4$). Circles, different mice. Lines, mean ± s.e.m. **f**, Flow cytometric analysis of mesenteric lymph node cells from the mice in **e**. Mean ± s.e.m.



| Genotype | Expected | Actual |
|---|---|---|
| *Casp8*$^{+/+}$ *Ripk1*$^{D138N/D138N}$ | 36.75 | 46 |
| *Casp8*$^{+/-}$ *Ripk1*$^{D138N/D138N}$ | 73.5 | 101 |
| *Casp8*$^{-/-}$ *Ripk1*$^{D138N/D138N}$ | 36.75 | 0 |

**Extended Data Fig. 4** | See next page for caption.

**Extended Data Fig. 4 | *Ripk1*$^{D325A/D325A}$ *Tnfr1*$^{-/-}$ and *Ripk1*$^{D138N, D325A/D138N, D325A}$ mice survive until after birth.**
**a**, *Ripk1*$^{D325A/+}$ *Tnfr1*$^{-/-}$ female with her 9-day-old male pups. Results representative of two *Ripk1*$^{D325A/D325A}$ *Tnfr1*$^{-/-}$ mice. **b**, Body weights of *Ripk1*$^{D138N,D325A/D138N,D325A}$ pups ($n = 3$ males, 2 females) and their littermates ($n = 1$ male, 4 females) aged 14–16 days. Circles, individual mice. Lines, mean ± s.e.m. $P = 3 \times 10^{-5}$; unpaired two-sided *t*-test with Welch's correction. **c**, Tissues from seven-day-old littermates stained with haematoxylin and eosin. Results representative of two *Ripk1*$^{D325A/D325A}$ *Tnfr1*$^{-/-}$ and three *Tnfr1*$^{-/-}$ mice. Scale bar, 50 μm. **d**, Tissues from 17-day-old *Ripk1*$^{D138N,D325A/D138N,D325A}$ mice. Asterisks, granulocytic infiltrates. Results representative of two mice. Scale bar, 100 μm. **e**, P4–P7 offspring from intercrossing *Casp8*$^{+/-}$ mice. **f**, E15.5 littermates. Results representative of two mice of each genotype. **g**, E13.5 embryo sections immunolabelled for p-RIPK3(T231/S232) (brown). Results representative of three *Casp8*$^{-/-}$ *Ripk1*$^{D138N/D138N}$ and two *Ripk1*$^{D138N/D138N}$ embryos. Scale bar, 200 μm. **h**, E12.5 littermates. Results representative of four *Casp8*$^{-/-}$ *Ripk1*$^{D325A/D325A}$ and three *Ripk1*$^{D325A/+}$ control embryos.
**i**, Western blots of primary MEFs. β-Actin loading control performed after IκBα. For gel source data, see Supplementary Fig. 1. Results representative of two independent experiments.



**Extended Data Fig. 5 | The RIPK3 scaffold contributes to TNF-induced apoptosis in *Ripk1*$^{D325A/D325A}$ MEFs.** **a**, **c**–**e**, Graphs show the percentage of primary MEFs (**a**, **c**) or BMDMs (**d**, **e**) that were viable at 24 h (**a**), 8 h (**c**, **e**), or 5 h (**d**) after treatment (−, untreated; T, TNF; N, Nec-1; i, TAK1 inhibitor 5Z-7-oxozeaenol; G, RIPK3 inhibitor GSK′843). Circles, cells from different mice (*Ripk1*$^{+/+}$, $n = 4$ (**a**), 2 (**c**), and 3 (**d**, **e**); *Ripk1*$^{D325A/D325A}$, $n = 4$ (**a**); *Ripk1*$^{D325A/D325A}$ *Ripk3*$^{-/-}$, $n = 3$ (**a**); *Ripk3*$^{-/-}$, $n = 4$ (**a**); *Ripk1*$^{D325A/+}$, $n = 2$ (**c**) and 3 (**d**, **e**)). Bars, mean ± s.e.m. Unpaired two-sided *t*-test. *not determined. **b**, **f**, Western blots of primary MEFs after 4 h treatment (**b**) or of BMDMs (**f**). β-Actin loading control performed after c-FLIP (**b**) or p-RIPK1 (**f**). For gel source data, see Supplementary Fig. 1. Results representative of two independent experiments.

nature research

Corresponding author(s): Kim Newton, Vishva M. Dixit

Last updated by author(s): 25 July 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Whole embryo images were acquired with Leica Application Suite v4.11.0. Immunohistochemistry and histology images were acquired with Leica Application Suite v4.6.0. Flow cytometry data were acquired with BD FACSDiva v8.0. Digital droplet PCR data were acquired with QuantaSoft 1.7.4.0917. Luminex data were acquired with Bio-Plex Manager 6.1.1. Caspase-Glo 8 data were acquired with PerkinElmer EnVision Manager 1.13.3009.1401. |
| Data analysis | Flow cytometry data were analysed with Flow Jo 10.3. Graphs were generated with Prism 7.0e. Digital droplet PCR data were analysed with QuantaSoft 1.7.4.0917. Luminex data were analysed with Bio-Plex Manager 6.1.1. Caspase-Glo 8 data were analysed with PerkinElmer EnVision Manager 1.13.3009.1401. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated during and/or analysed during the current study are available from the corresponding authors on reasonable request.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | No sample size calculations were performed. Whenever possible, tissues and cells from at least 3 animals per genotype were analysed to be sure differences were reproducible. Variability in the ex vivo assays used in this study tends to be low, so n=3 is the accepted norm in the field. Larger numbers (n=5-10 per group) were used for the in vivo TNF challenge studies to account for the greater variability between wild-type controls in these experiments. |
| Data exclusions | No data were excluded. |
| Replication | Whenever possible, readouts were performed with at least 3 animals of a given genotype and all attempts at replication were successful. |
| Randomization | Mice were grouped according to genotype, not randomized. Where possible, animals were age- and sex-matched. |
| Blinding | Embryos were imaged blindly before genotypes were known.<br>However, adult mice were selected and treated by the same individual, so blinding was not possible. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | Antibodies |
| ☐ | ☒ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology |
| ☐ | ☒ | Animals and other organisms |
| ☒ | ☐ | Human research participants |
| ☒ | ☐ | Clinical data |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☒ | ☐ | ChIP-seq |
| ☐ | ☒ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

## Antibodies

Antibodies used

Antibodies used for flow cytometry were from BD Biosciences (145-2C11 anti-CD3, cat#557596, lot#7201885, or cat#553061, lot#41936; RA3-6B2 anti-B220, cat#553088, lot#3352893, or cat#553092, lot#50613; 2.4G2 anti-CD16/CD32, cat#553142, lot#M063276).

Antibodies used for western blots were from Genentech (1G6 anti-mouse RIPK3; GEN135-35-9 anti-mouse phospho-RIPK3 Thr231, Ser232; 10C7 anti-mouse RIPK1; GEN150-33-4 anti-mouse phospho-RIPK1 Ser166, Thr169; 1.28E12 anti-mouse FADD; 2.21H2 anti-mouse c-FLIP; GN55-9 anti-mouse CYLD), Enzo Life Sciences (1G12 anti-mouse caspase-8, cat#ALX-804-447-C100, lot#11131404), BD Biosciences (38/RIP anti-RIPK1, cat#610459, lot#5177938), MP Biomedicals (C4 anti-beta actin, cat#69100, lot#QR14180), Millipore (3H1 anti-MLKL, cat#MABC604, lot#2677923), Abcam (EPR9515(2) anti-phospho-MLKL S345, cat#ab196436 , lot#GR3204546-5), Cell Signaling Technology (p-ERK, cat#9101, lot#28; ERK, cat#9102, lot#26; p-JNK, cat#4668, lot#12; JNK, cat#9258, lot#6; p-IkappaBalpha, cat#2859, lot#13; IkappaBalpha, cat#9242, lot#9; p-p65 cat#3033, lot#16; p65, cat#8242, lot#8; p-p38, cat#9211, lot#23; p38, cat#8690, lot#4; caspase-3, cat#9662, lot#18; cleaved caspase-3, cat#9664, lot#21; cleaved caspase-8, cat#8592, lot#3), Sigma (M2 anti-Flag, cat#A8592, lot#SLBD9930), Genetex (anti-myc polyclonal, cat#GTX21261, lot#821805083), and Jackson Immunoresearch (HRP-anti-mouse, cat#115-035-174; lot#139280; HRP-anti-rabbit, cat#211-032-171; lot#120918; HRP-anti-rat, cat#112-035-175; lot#115169).

Antibodies used for immunofluorescence were from BD Biosciences (MEC 13.3 anti-PECAM1, cat #550274, lot#2243973), Cell Signaling Technology (anti-cleaved caspase-3, cat #9661, lot#38), and Jackson Immunoresearch (Cy3-anti-rabbit, cat#711-165-152, lot#86183; Cy5-anti-rat, cat#112-175-167, lot#123967).

Antibodies for immunohistochemistry were from Genentech (GEN135-35-9 anti-mouse phospho-RIPK3 Thr231, Ser232; 10C7

| | |
|---|---|
| | anti-mouse RIPK1), Enzo Life Sciences (1G12 anti-mouse caspase-8, cat#ALX-804-447-C100, lot#11131404), and Vector labs (rabbit anti-rat cat#AI-4001, lot#Y0920). |
| Validation | 1G6 anti-mouse RIPK3 is validated for WB using RIPK3 KO mouse tissues in Newton et al (2004) Mol. Cell. Biol. 24:1464-1469, and for IHC using RIPK3 KO mouse tissues in Newton et al (2014) Science 343:1357-1360.<br>GEN135-35-9 anti-mouse phospho-RIPK3 T231, S232 is validated for WB and IHC in Newton et al (2016) Nature 540:129-133.<br>10C7 anti-mouse RIPK1 is validated for IHC using RIPK1 KO mouse tissues in Newton et al (2014) Science 343:1357-1360. We have also validated its use for WB using RIPK1 KO mouse tissues.<br>38/RIP anti-RIPK1 is validated for WB using RIPK1 KO mouse cells in Newton et al (2016) Nature 540:129-133.<br>GEN150-33-4 anti-mouse phospho-RIPK1 S166, T169 is validated for WB in Heger et al (2018) Nature 559:120-124.<br>1.28E12 anti-mouse FADD was validated for WB using cells from Fadd-/- Mlkl-/- mice.<br>2.21H2 anti-mouse c-FLIP was validated for WB using cells from 3xFlag-c-FLIP knock-in mice.<br>GN55-9 anti-mouse CYLD was validated for WB using cells from Cyld-/- mice.<br>1G12 anti-mouse casp8 is validated for WB in O'Reilly et al (2004) Cell Death Differ 11:724–736, and we have confirmed this using Casp8-/- Mlkl-/- mouse cells (Extended data fig. 1d). It was validated for IHC using Casp8-/- mouse embryos (Extended data fig. 3c).<br>3H1 anti-MLKL is validated for WB using Mlkl-/- mouse cells in Murphy et al (2013) Immunity 39:443-453.<br>Validation data for commercial antibodies are available on vendor websites. |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | 293T cells (ATCC CRL-3216). |
| Authentication | Cells not authenticated. |
| Mycoplasma contamination | Cells negative for mycoplasma. |
| Commonly misidentified lines (See ICLAC register) | Not used. |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | All mice (Mus musculus) were maintained on a C57BL/6N genetic background. Mutant strains included Ripk1 D138N/D138N (Newton et al. 2014 Science 343:1357-1360), Ripk1 D325A/+ (this study), Ripk3-/- (Newton et al. 2014 Science 343:1357-1360), Casp8+/- (Newton et al. 2014 Science 343:1357-1360), Casp8 1xDA/1xDA (this study), Casp8 3xDA/3xDA (this study), Casp8 4xDA/4xDA (this study), Cyld 1xDA/1xDA (this study), Fadd+/- (this study), Cflar 2xDA/2xDA (this study), Mlkl-/- (Murphy et al. 2013 Immunity 39:443-453), and Tnfr1-/- (Peschon et al. 1998 J immunol 160:943-952). Mice of both sexes were analysed, ranging in age from E10.5 through 21 months. |
| Wild animals | This study did not involve wild animals. |
| Field-collected samples | This study did not involve samples collected from the field. |
| Ethics oversight | All mouse studies were approved by the Genentech institutional animal care and use committee (IACUC). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Flow Cytometry

### Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

| | |
|---|---|
| Sample preparation | Thymocyte or lymph node single cell suspensions were prepared by gently pressing the thymus or mesenteric lymph node through a 40 um cell strainer with the rubber plunger from a 3 ml syringe. An aliquot of cells was stained with acridine orange and DAPI (solution 18; Chemometec) and the viable cell concentration determined using a Nucleoview NC-250 cell counter (Chemometec). Adherent MEFs or BMDMs in culture were trypsinized and pooled with floating dead cells, then stained with propidium iodide. |

| | |
|---|---|
| Instrument | BD FACSCantoII (BD Biosciences) |
| Software | Data was acquired using BD FACSDiva software v8.0, and analysed using FlowJo 10.3. |
| Cell population abundance | No sorting was performed. |
| Gating strategy | Leukocytes were identified based on their forward scatter (FSC-A) and side scatter (SSC-A) profiles. Dead cells that stained with propidium iodide or 7-AAD (BD Biosciences), plus doublets, identified by their FSC-A versus FSC-W profiles, were excluded from analyses of cells stained for CD3 and B220. Survival studies only excluded cell debris with a small FSC, that was well separated from cells. |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

https://doi.org/10.1038/s41586-019-1646-9

# Spliceosomal disruption of the non-canonical BAF complex in cancer

Daichi Inoue[1,2,10], Guo-Liang Chew[3,4,10], Bo Liu[1], Brittany C. Michel[5,6,7], Joseph Pangallo[3,4], Andrew R. D'Avino[6,7], Tyler Hitchman[1], Khrystyna North[3,4,8], Stanley Chun-Wei Lee[1], Lillian Bitner[1], Ariele Block[1], Amanda R. Moore[1], Akihide Yoshimi[1], Luisa Escobar-Hoyos[1], Hana Cho[1], Alex Penson[1], Sydney X. Lu[1], Justin Taylor[1], Yu Chen[1,9], Cigall Kadoch[6,7], Omar Abdel-Wahab[1,9]* & Robert K. Bradley[3,4,8]*

**SF3B1 is the most commonly mutated RNA splicing factor in cancer[1–4], but the mechanisms by which *SF3B1* mutations promote malignancy are poorly understood. Here we integrated pan-cancer splicing analyses with a positive-enrichment CRISPR screen to prioritize splicing alterations that promote tumorigenesis. We report that diverse *SF3B1* mutations converge on repression of BRD9, which is a core component of the recently described non-canonical BAF chromatin-remodelling complex that also contains GLTSCR1 and GLTSCR1L[5–7]. Mutant SF3B1 recognizes an aberrant, deep intronic branchpoint within *BRD9* and thereby induces the inclusion of a poison exon that is derived from an endogenous retroviral element and subsequent degradation of *BRD9* mRNA. Depletion of BRD9 causes the loss of non-canonical BAF at CTCF-associated loci and promotes melanomagenesis. BRD9 is a potent tumour suppressor in uveal melanoma, such that correcting mis-splicing of *BRD9* in *SF3B1*-mutant cells using antisense oligonucleotides or CRISPR-directed mutagenesis suppresses tumour growth. Our results implicate the disruption of non-canonical BAF in the diverse cancer types that carry *SF3B1* mutations and suggest a mechanism-based therapeutic approach for treating these malignancies.**

*SF3B1* is subject to recurrent missense mutations at specific residues in myeloid leukaemia[1,2] and lymphoid leukaemia[3,8] as well as in solid tumours, at rates of up to 14–29% for uveal melanoma (UVM)[9–12] and 65–83% for myelodysplastic syndromes with ring sideroblasts[1,2]. Consistent with the critical role of SF3B1 in the recognition of 3′ splice sites[13], several previous studies have reported that *SF3B1* mutations induce widespread usage of abnormal 3′ splice sites[10,14,15]. Although many mis-spliced genes have been identified in *SF3B1*-mutant samples, few of these have been functionally implicated in driving disease.

We hypothesized that effectors of the pro-tumorigenic consequences of *SF3B1* mutations might appear as pan-cancer targets of mutant SF3B1. We accordingly identified mis-spliced events that were shared between erythroleukaemic (K562) and UVM (MEL270) cells that expressed wild-type *SF3B1* or the most-common *SF3B1* mutation, $SF3B1^{K700E}$. A compact set of 40 events exhibited concordant splicing changes, and was sufficient to infer *SF3B1* mutational status across 249 samples from patients with chronic lymphocytic leukaemia, myelodysplastic syndromes and UVM (Fig. 1a, Extended Data Fig. 1a, Supplementary Tables 1–3).

We designed a single-guide RNA (sgRNA) library that targeted both pan-cancer and cancer-type-specific targets of mutant SF3B1, focusing on genes for which *SF3B1* mutations are predicted to cause mis-splicing that triggers nonsense-mediated RNA decay (NMD) (Fig. 1b, Supplementary Table 4). We tested whether the knockout of any of these genes promoted the transformation of Ba/F3 cells (a mouse cell line with a wild-type spliceosome, with a requirement for IL-3 that can be overcome by oncogenic lesions) (Fig. 1c). In addition to the positive control *Pten*, our screen revealed that the loss of *Brd9* promoted the transformation of Ba/F3 cells (Fig. 1d, Extended Data Fig. 1b–d, Supplementary Tables 5, 6). *Brd9* was a notable hit because *BRD9* exhibited notable mis-splicing in all cohorts of patients with *SF3B1*-mutant cancer (Fig. 1e). *Brd9* knockout conferred cytokine independence to mouse 32Dcl3 cells, and growth advantage to human cancer cells with a wild-type spliceosome derived from UVM, cutaneous melanoma, and pancreatic cancer (Extended Data Fig. 1d–f). By contrast, acute myeloid leukaemia cells with rearranged *MLL* (also known as *KMT2A*) required BRD9 for growth (Extended Data Fig. 1g), as previously reported[16].

*SF3B1* mutations cause the exonization of a *BRD9* intronic sequence, which results in the inclusion of a poison exon that interrupts the open reading frame of *BRD9*. This *BRD9* poison exon is derived from a primate-specific endogenous retroviral element, explaining its absence from mice (Extended Data Fig. 1h, i). We confirmed that the inclusion of the poison exon was induced by the expression of endogenous or ectopic mutant SF3B1 in K562 and NALM-6 cells, whereas *SF3B1* knockdown in *SF3B1* wild-type cells had no effect (Extended Data Fig. 1j–m). The poison exon was included in an *SF3B1*-mutation-dependent manner in diverse cell lines and in samples of chronic lymphocytic leukaemia, myelodysplastic syndromes and UVM that bear 19 different *SF3B1* mutations—but not in healthy tissues (Extended Data Fig. 1m–p, Supplementary Table 7).

The inclusion of the *BRD9* poison exon triggered NMD and reduced the half-life of *BRD9* mRNA and steady-state levels of full-length BRD9 protein (Extended Data Fig. 1q–w). Patients with *SF3B1* mutations exhibited reduced total levels of *BRD9* mRNA relative to patients with wild-type *SF3B1* (Extended Data Fig. 1x). We tested whether the inclusion of the poison exon could result in the production of C-terminally truncated BRD9 by knocking an N-terminal haemagglutinin tag into the *BRD9* locus in MEL270 and K562 cells that transgenically express wild-type or mutant SF3B1 (Extended Data Fig. 2a–c). Mutant SF3B1 suppressed levels of full-length BRD9 protein, without generating a truncated BRD9 protein (Fig. 1f).

*SF3B1* mutations promote the use of cryptic 3′ splice sites[10,14,15], probably by altering the normal role of SF3B1 in branchpoint recognition[17]. We therefore mapped the *BRD9* branchpoints used in K562, MEL270 and T47D (breast cancer) cells that express mutant SF3B1 (Fig. 2a, Extended Data Fig. 2d–f). The inclusion of the poison exon was associated with an unusually close branchpoint (close branchpoints are rare and normally inefficiently recognized[18]). Mutating the aberrant branchpoint abolished poison exon recognition (Fig. 2b, Extended Data

---

[1] Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2] Department of Hematology-Oncology, Institute of Biomedical Research and Innovation, Foundation for Biomedical Research and Innovation at Kobe, Kobe, Japan. [3] Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [4] Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [5] Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [6] Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7] Biomedical and Biological Sciences Program, Harvard Medical School, Boston, MA, USA. [8] Department of Genome Sciences, University of Washington, Seattle, WA, USA. [9] Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [10] These authors contributed equally: Daichi Inoue, Guo-Liang Chew. *e-mail: abdelwao@mskcc.org; rbradley@fredhutch.org

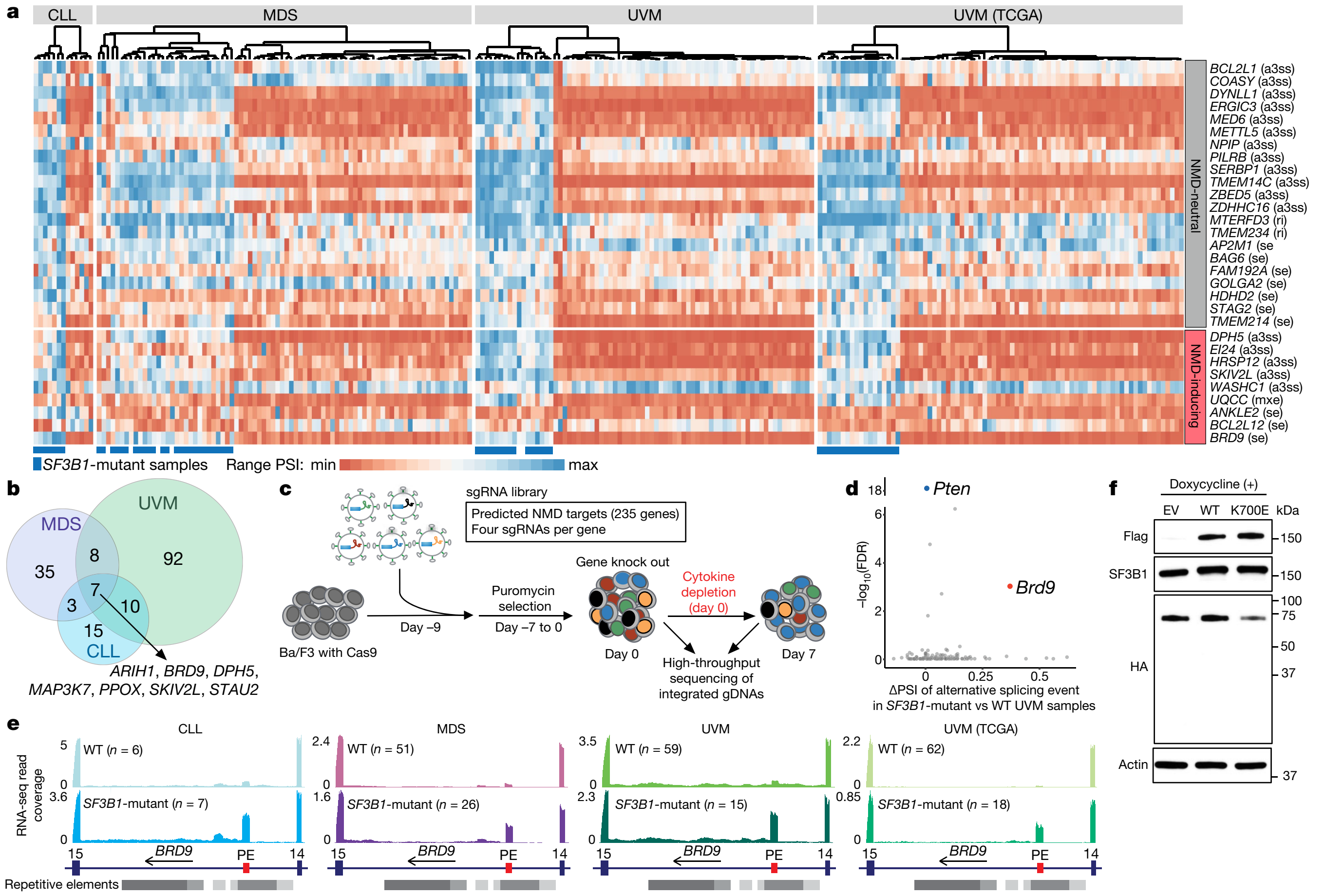


**Fig. 1 | *BRD9* mis-splicing causes BRD9 loss and proliferative advantage in *SF3B1*-mutated cancers. a**, Unsupervised clustering of patient samples on the basis of events that are differentially spliced in UVM (MEL270) and myeloid leukaemia (K562) cells that express *SF3B1*$^{K700E}$ versus wild-type (WT) *SF3B1*. a3ss, alternative 3′ splice site; CLL, chronic lymphocytic leukaemia (data are from ref. [15]); MDS, myelodysplastic syndromes (data are from ref. [27]); mxe, mutually exclusive exons; PSI, percentage spliced in (fraction of mRNA that corresponds to the mutant SF3B1-promoted isoform), with per-event and per-cohort range normalization; ri, retained intron; se, skipped exon; TCGA, The Cancer Genome Atlas. Data for UVM are from ref. [10] (middle) or the TCGA (right). *FAM192A*, *HRSP12*, *MTERFD3*, *NPIP* and *UQCC* are also known as *PSME3IP1*, *RIDA*, *MTERF2*, *NPIPA1* and *UQCC1*, respectively. **b**, Genes for which mutant SF3B1 promotes an isoform predicted to trigger NMD (alternative 3′ splice site and skipped exon events only) in one or more cohorts. **c**, CRISPR–Cas9-based positive-selection screen targeting genes for which mutant SF3B1 promotes an isoform predicted to trigger NMD. **d**, Per-gene scatter plot comparing CRISPR screen enrichment (*y* axis) to differential splicing in TCGA cohort of patients with UVM (*x* axis). *Pten* was used as a positive control. $n = 6$ biologically independent experiments. Per-gene significance computed with two-sided correlation-adjusted mean rank gene set (CAMERA) test. The false-discovery rate (FDR) was computed using the Benjamini–Hochberg method. **e**, *BRD9* RNA sequencing (RNA-seq) read coverage in patient samples. *n*, number of patients. PE, *BRD9* poison exon; 14 and 15, flanking constitutive exons. Repetitive elements from RepeatMasker[28]. **f**, Western blot for N-terminally haemagglutinin (HA)-tagged endogenous BRD9 in MEL270 cells transduced with empty vector (EV) or doxycycline-inducible Flag–SF3B1(WT) or Flag–SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments.

Fig. 2g). Consistent with the lack of an obvious polypyrimidine tract upstream of the poison exon, neither *U2AF1* nor *U2AF2* knockdown compromised poison exon recognition, whereas introducing a poly(Y) tract resulted in robust inclusion of the poison exon even in wild-type cells (Fig. 2b, Extended Data Fig. 2h–j). Finally, we identified a putative exonic splicing enhancer that was essential for inclusion of the poison exon (Fig. 2c, Extended Data Fig. 2k). We confirmed that the aberrant branchpoint, lack of a polypyrimidine tract, and exonic splicing enhancer were essential for poison exon recognition in the context of *SF3B1*$^{R625H}$, the most common mutation of *SF3B1* in UVM (Extended Data Fig. 2l–n). Disrupting the 3′ splice site and/or exonic splicing enhancer of the poison exon with CRISPR-directed mutagenesis markedly increased the levels of BRD9 protein in UVM cells with mutated *SF3B1* (Fig. 2d, Extended Data Fig. 2o, Supplementary Table 8), but had no effect on *BRD9* splicing or expression in cells with wild-type *SF3B1* (Extended Data Fig. 2p–r).

Several studies have recently described BRD9 as part of a non-canonical (nc) BAF complex, which is biochemically distinct from canonical BAF and polybromo-associated BAF[5–7] (Fig. 3a). Although ncBAF is not recurrently mutated in cancer—unlike canonical BAF and polybromo-associated BAF (Extended Data Fig. 3a)—our data suggested that ncBAF is nonetheless frequently disrupted via *SF3B1* mutations.

We investigated the consequences of BRD9 loss by *SF3B1* mutations for ncBAF function. Immunoprecipitation and mass spectrometry to identify the chromatin-associated interaction partners of BRD9 in K562 cells specifically recovered ncBAF components (Extended Data Fig. 3b, c, Supplementary Table 9). We confirmed these results by immunoblotting against shared and complex-specific components of canonical BAF, polybromo-associated BAF and ncBAF in K562 and UVM cells (Fig. 3b, Extended Data Fig. 3d). Expression of mutant, but not wild-type, SF3B1 reduced the levels of BRD9 protein and abolished interactions between BRG1 and GLTSCR1 while leaving interactions between BRG1 and BAF155 intact, which indicates that *SF3B1* mutations specifically perturb ncBAF rather than disrupting all BAF complexes (Fig. 3c, Extended Data Fig. 3e). Chemical degradation of BRD9[19] or *BRD9* knockout similarly reduced the BRG1–GLTSCR1 interaction (Fig. 3c, Extended Data Fig. 3f). We next identified the BRD9 domains that are necessary for ncBAF formation by generating 3×Flag–BRD9 deletion mutants

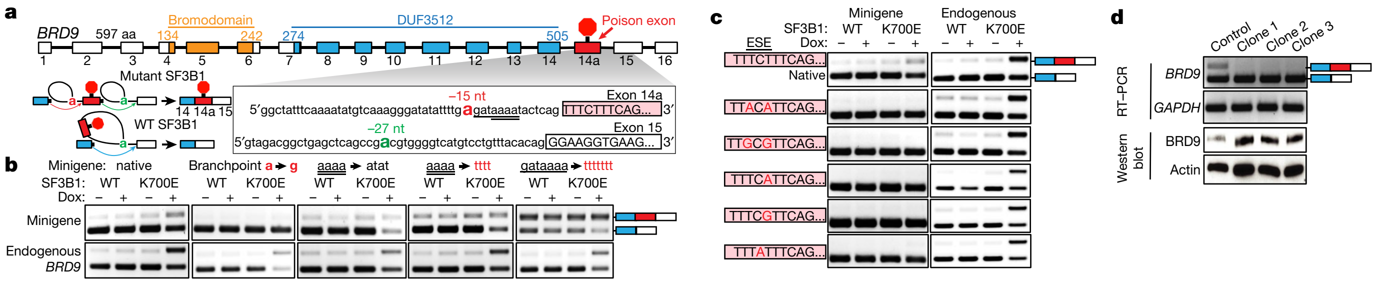


**Fig. 2 | Mutant SF3B1 recognizes an aberrant, deep intronic branchpoint within *BRD9*. a**, *BRD9* gene structure and protein domains. Inset illustrates the branchpoints used when the poison exon is included (top) or excluded (bottom). Single and double underlining indicates sequence motifs that were subsequently mutated. aa, amino acid. **b**, PCR with reverse transcription (RT–PCR) analysis of inclusion of the *BRD9* poison exon in a minigene (top) or endogenous (bottom) context, following transfection of minigenes with the illustrated mutations into MEL270 cells with doxycycline (dox)-inducible Flag–SF3B1(WT) or Flag–SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments. Native, no mutations. **c**, As in **b**, but for minigene mutations (shown in red) at the 5′ end of the poison exon. ESE, exonic splicing enhancer. **d**, RT–PCR (top) illustrating the loss of inclusion of the *BRD9* poison exon, and corresponding western blot (bottom) in MEL202 (*SF3B1*$^{R625G}$) clones following CRISPR–Cas9 targeting of the poison exon. Indels are illustrated in Extended Data Fig. 2o. Control, unedited cells. Representative images from $n = 2$ (RT–PCR) and $n = 3$ (western blot) biologically independent experiments.

and testing for interactions with GLTSCR1 and GLTSCR1L. These experiments revealed that the DUF3512 domain of BRD9 mediates its interactions with GLTSCR1 and GLTSCR1L (Extended Data Fig. 3g–h).

We next determined how *SF3B1* mutations altered ncBAF localization to chromatin. We mapped the genome-wide binding of the pan-BAF component BRG1, and the ncBAF-specific components BRD9 and GLTSCR1, in MEL270 cells that express wild-type or mutant SF3B1. We additionally performed the same chromatin immunoprecipitation with sequencing (ChIP–seq) experiments after treatment with dimethylsulfoxide (DMSO) or a BRD9 degrader to identify BRD9-dependent effects. BRD9 and GLTSCR1 exhibited substantial co-localization, consistent with their mutual requirement for ncBAF formation, and were found at a subset of the loci bound by BRG1 (Fig. 3d). BRD9 and GLTSCR1 bound to promoters, gene bodies, and probable enhancers, with focal binding at promoters relative to BRG1 (Fig. 3e, Extended Data Fig. 4a). CTCF motifs exhibited notable co-localization with GLTSCR1, but only modest co-localization with BRG1 (Fig. 3f, Extended Data Fig. 4b).

We then tested how the depletion of BRD9, induced by *SF3B1*$^{K700E}$ or by chemical degradation of BRD9, altered ncBAF localization. We defined the genomic loci bound by GLTSCR1 in all samples as constitutive sites. Conversely, we defined genomic loci bound by GLTSCR1 in both control (wild-type SF3B1 or DMSO) but not BRD9-depleted

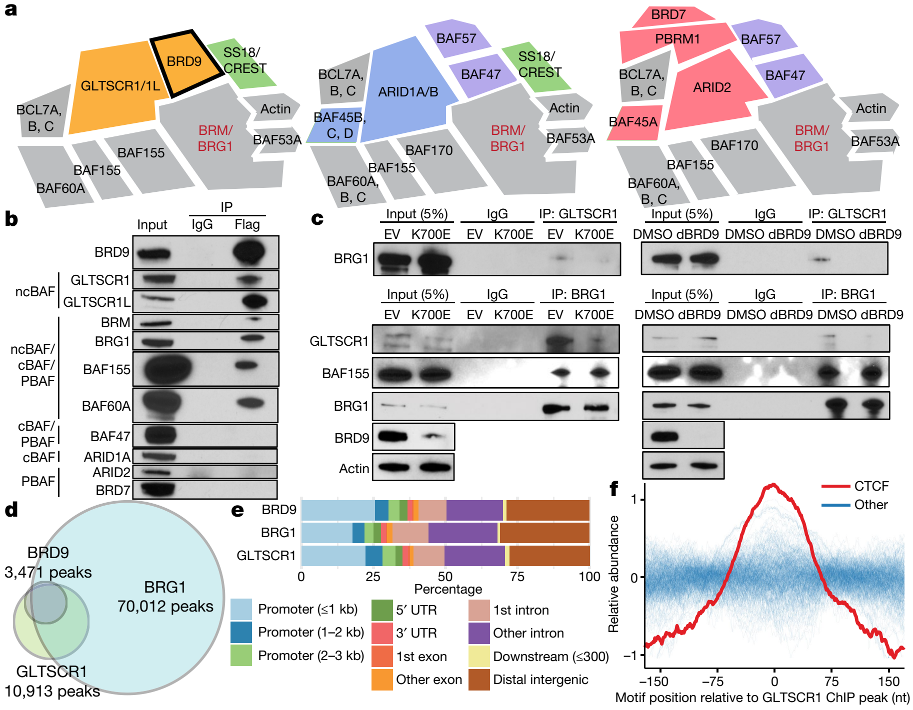


**Fig. 3 | BRD9 loss perturbs the formation and localization of the ncBAF complex. a**, Schematic of non-canonical BAF (ncBAF; left), canonical BAF (cBAF; middle) and polybromo-associated BAF (PBAF; right) complexes[5,6]. Solidus denotes one of the proteins is present; comma denotes one or more of the proteins are present or that mutually exclusive inclusion of proteins may occur. **b**, Cross-linking and immunoprecipitation (IP) with IgG or Flag followed by immunoblotting in K562 cells that express 3×Flag–BRD9. Representative images from $n = 3$ biologically independent experiments. **c**, Immunoprecipitation with GLTSCR1 or BRG1 antibody followed by immunoblotting in MEL270 cells that express exogenous SF3B1(K700E) (left) or were treated with BRD9 degrader[19] (dBRD9) (right). Representative images from $n = 3$ biologically independent experiments. **d**, Overlap of consensus BRD9, BRG1 and GLTSCR1 ChIP–seq peaks called in both MEL270 control samples (DMSO and ectopic expression of wild-type SF3B1). **e**, Genomic localization of BRD9-, BRG1-, and GLTSCR1-bound loci in **d**. UTR, untranslated region. **f**, Distributions of transcription factor-binding motifs at GLTSCR1-bound loci (20 nucleotide (nt) rolling mean). $n = 401$ transcription factors analysed.

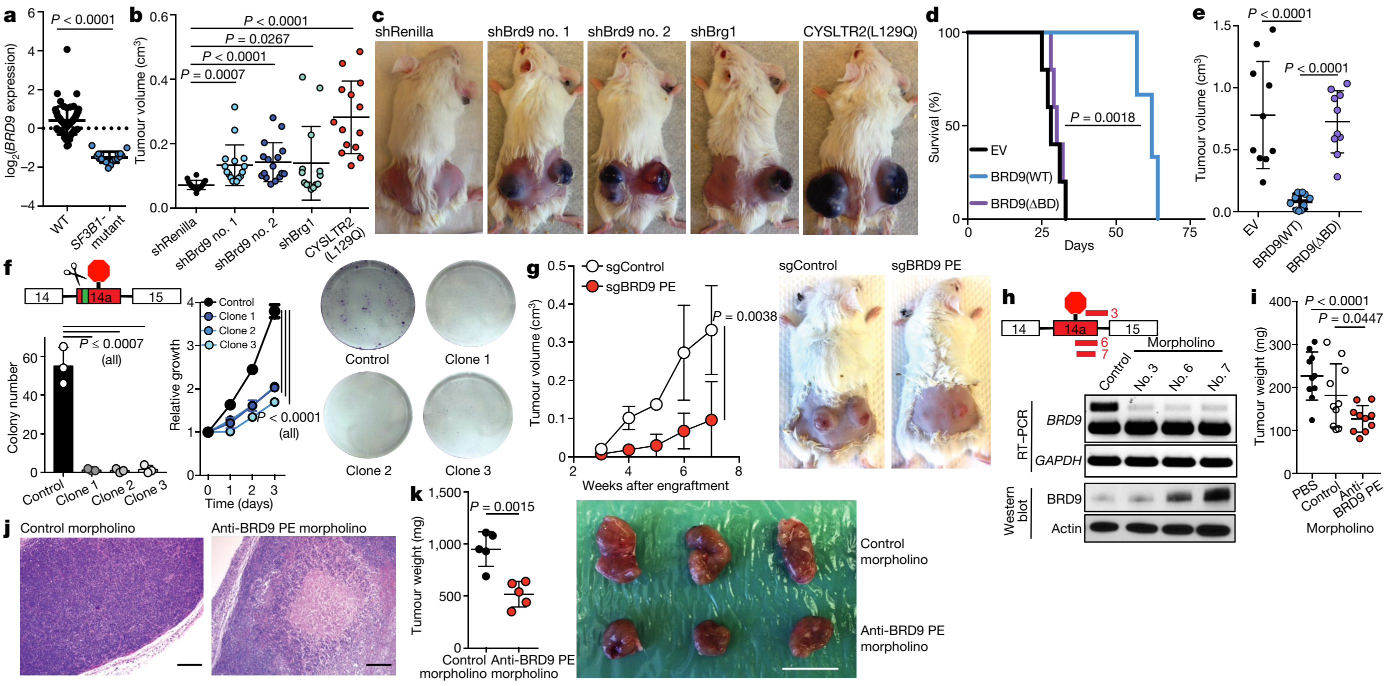


**Fig. 4 | BRD9 is a therapeutically targetable tumour suppressor in melanoma. a**, *BRD9* expression (*z*-score normalized) in TCGA UVM samples with ($n = 18$) or without ($n = 62$) *SF3B1* mutations. *P* value calculated by two-sided *t*-test. **b**, Tumour volume 49 days after subcutaneous engraftment of Melan-a cells transduced with the indicated shRNAs into SCID mice. $n = 16$, 16, 16, 14 and 14 tumours per group (left to right). Error bars, mean ± s.d. *P* values calculated by two-sided *t*-test. **c**, Representative mice from **b** at day 63. **d**, Survival of SCID mice engrafted with MEL270 cells that express empty vector, full-length wild-type BRD9 or a BRD9 bromodomain-deletion mutant (ΔBD). $n = 5$ mice per group. *P* value calculated by log-rank test. **e**, Tumour volume from experiments shown in **d**, 21 days after engraftment. $n = 10$ tumours per group. Error bars, mean ± s.d. *P* values calculated by two-sided *t*-test. **f**, Colony number (left) and representative images (right) of MEL202 cells (*SF3B1*$^{R625G}$) without (control) or with (clone 1, clone 2 and clone 3) CRISPR–Cas9-induced disruption of the *BRD9* poison exon. Indels are illustrated in Extended Data Fig. 2o. $n = 3$ biologically independent experiments. Error bars, mean ± s.d. *P* values calculated by two-sided *t*-test at day 3 (middle). **g**, Tumour volume (left) and representative images (right) of mice engrafted with control or clone 1 cells from **f**. $n = 6$ tumours per group. Error bars, mean ± s.d. *P* value calculated by two-sided *t*-test at week 7. **h**, ASO design (top), RT–PCR (middle) and western blot (bottom) for BRD9. MEL202 cells (*SF3B1*$^{R625G}$) were treated with a non-targeting (control) or targeting morpholino at 10 μM for 24 h. Representative images from $n = 3$ biologically independent experiments. **i**, Tumour weight following 16 days of in vivo treatment of MEL202-derived xenografts (*SF3B1*$^{R625G}$) with PBS or a non-targeting (control) or poison-exon-targeting (no. 6) morpholino (12.5 mg kg$^{-1}$, every other day to a total of 8 intratumoral injections). $n = 10$ tumours per group. Error bars, mean ± s.d. *P* values calculated by two-sided *t*-test. **j**, Haematoxylin and eosin (H&E) images of tumours from **i**. Scale bars, 200 μm. Representative images from $n = 3$ biologically independent histological analyses. **k**, Tumour weight (left) and representative images (right) following in vivo morpholino treatment of a patient-derived rectal melanoma xenograft (*SF3B1*$^{R625C}$). Scale bar, 1 cm. $n = 5$ tumours per group. *P* value calculated by two-sided *t*-test. Error bars, mean ± s.d.

(mutant SF3B1 or BRD9 degradation) samples as BRD9-sensitive sites (Extended Data Fig. 4c). GLTSCR1 peaks were more sensitive to BRD9 loss than were BRG1 peaks, and CTCF motifs were uniquely enriched in BRD9-sensitive loci ($P < 1 \times 10^{-8}$) versus constitutive GLTSCR1-bound loci (Extended Data Fig. 4d). CTCF was similarly highly enriched at BRG1-bound loci that were BRD9-sensitive ($P < 1 \times 10^{-55}$) (Extended Data Fig. 4e, f). We conclude that the BRD9 loss induced by *SF3B1* mutations causes specific loss of ncBAF at CTCF-associated loci.

We identified genes with BRD9-sensitive ncBAF binding in their promoters or enhancers and found that BRD9 loss in UVM preferentially affects genes involved in apoptosis and cell growth, adhesion and migration (Extended Data Fig. 4g). To understand how BRD9 loss altered gene expression, we identified genes with promoters that exhibited BRD9-sensitive ncBAF binding and that were differentially expressed in patients with UVM with mutant versus wild-type *SF3B1*. Loss of ncBAF binding was associated with promotion as well as repression of gene expression, suggesting that ncBAF—similar to other SWI/SNF complexes—has both activating and repressive roles[20] (Extended Data Fig. 4h–j).

Several recent studies have reported that BRD9 is required for the survival of some cancer types, particularly cancers with mutations that affect polybromo-associated BAF and canonical BAF[6,16,21]. Because *BRD9* loss conferred a proliferative advantage to types of cancer with recurrent *SF3B1* mutations (Fig. 1d, Extended Data Fig. 1), we wondered whether normalizing the levels of BRD9 might suppress the growth of *SF3B1*-mutant cells.

As *SF3B1* is recurrently mutated in uveal (Fig. 4a), mucosal and cutaneous melanomas, we first tested whether *BRD9* loss induced melanomagenesis in vivo. We transduced non-tumorigenic mouse melanocytes (Melan-a cells), which require oncoprotein expression for sustained growth, with a non-targeting short hairpin RNA (shRNA) (Extended Data Fig. 5a), doxycycline-inducible shRNAs targeting *Brd9* or *Brg1* (also known as *Smarca4*) or a cDNA encoding the oncoprotein CYSLTR2(L129Q) (as a positive control[22]). Knockdown of either *Brd9* or *Brg1* resulted in potent tumour growth, augmented melanocyte pigmentation, and expression of melanocyte-lineage-specific genes in vivo (Fig. 4b, c, Extended Data Fig. 5b–g).

We next tested whether *Brd9* expression influences metastasis. *Brd9* knockdown significantly increased the number of pulmonary metastatic foci following intravenous injection of cells from a mouse model of melanoma (B16) or of human UVM (92.1) cells into mice (Extended Data Fig. 6a–f). By contrast, restoring *Brd9* expression in established tumours in vivo, by withdrawing doxycycline, suppressed tumour growth (Extended Data Fig. 6g, h). Similarly, ectopic expression of full-length BRD9, but not the bromodomain- or DUF3512-deletion mutants, suppressed the growth of UVM cell lines and xenografts (Fig. 4d, e, Extended Data Fig. 6i–k). These data demonstrate that loss of *Brd9* promotes cell transformation, tumour maintenance, and metastatic progression, and that the bromodomain and DUF3512 domain of BRD9 are essential for its anti-proliferative effects.

We sought to understand how BRD9 loss promotes melanoma tumorigenesis. We identified BRD9-bound genes that exhibited dysregulated expression in samples from patients with UVM with mutated versus wild-type *SF3B1*, and in isogenic UVM cells with or without mutant SF3B1 and with or without forced loss of BRD9. *HTRA1*, a known tumour suppressor in melanoma[23,24], was the most downregulated gene in UVM (Extended Data Fig. 7a–c). HTRA1 was suppressed by mutant SF3B1 expression and BRD9-degradation treatment of UVM cells with wild-type *SF3B1*, and mutagenesis of the *BRD9* poison exon increased levels of HTRA1 in UVM cells with mutated *SF3B1* (Extended Data Fig. 7d, e). *HTRA1* is bound by ncBAF in UVM, and this binding is reduced by mutant SF3B1 (Extended Data Fig. 7f). *HTRA1* knockdown promoted the growth of UVM cells with wild-type *SF3B1*, and ectopic expression of *HTRA1* suppressed the growth of UVM cells with mutated *SF3B1* (Extended Data Fig. 7g–k). These data suggest that perturbation of ncBAF-dependent regulation of *HTRA1* contributes to the pro-tumorigenic effects of BRD9 loss.

We next tested whether correcting *BRD9* mis-splicing suppressed tumorigenesis. CRISPR-based mutagenesis of the poison exon markedly slowed the growth of cells with mutated *SF3B1*, but not of wild-type cells, both in vitro and in vivo (Figs. 2d, 4f, g, Extended Data Figs. 2o–r, 8). We then designed antisense oligonucleotides (ASOs) to block the inclusion of the *BRD9* poison exon (Fig. 4h, Extended Data Fig. 9a). We treated *SF3B1*-mutated cells with a non-targeting (control) or poison-exon-targeting ASO, and measured *BRD9* splicing, BRD9 protein levels, and cell growth. Each targeting ASO prevented the inclusion of the poison exon, increased the level of BRD9 protein, and suppressed cell growth relative to the control ASO (Fig. 4h, Extended Data Fig. 9b). The relative abilities of each ASO to restore BRD9 protein levels and suppress cell growth were strongly correlated, consistent with on-target effects.

We therefore tested whether ASO treatment slowed tumour growth in vivo. We treated *SF3B1*-mutated xenografts (derived from MEL202 cells) with each ASO via intratumoral injection for 16 days. Treatment with the poison-exon-targeting ASO—but not with the non-targeting ASO—corrected *BRD9* mis-splicing, significantly reduced tumour growth, and induced tumour necrosis (Fig. 4i, j, Extended Data Fig. 9c–f). We observed a similar ASO efficacy in a patient-derived xenograft model of rectal melanoma with the $SF3B1^{R625C}$ mutation (Fig. 4k, Extended Data Fig. 9g–i, Supplementary Table 7). By contrast, when we performed an identical experiment with a patient-derived xenograft model of UVM that lacked an *SF3B1* mutation, treatment with the poison-exon-targeting ASO had no effect (Extended Data Fig. 9j–l, Supplementary Table 7). We conclude that correcting *BRD9* mis-splicing restores the tumour suppressor activity of BRD9 in cancers with *SF3B1* mutations.

Although recognition of the *BRD9* poison exon requires mutant SF3B1, *BRD9* mis-splicing and ncBAF disruption may also have roles in cancers with wild-type *SF3B1*. We identified significant pan-cancer expression correlations between *BRD9* and many genes that encode RNA-binding proteins, as well as six additional *BRD9* isoforms that are predicted to trigger NMD that are expressed in cancers with wild-type *SF3B1* and are predictive of *BRD9* expression (Extended Data Fig. 10, Supplementary Table 10). A recent study has also identified a promoter polymorphism associated with decreased *GLTSCR1* (also known as *BICRA*) expression as a common risk allele for acute myeloid leukaemia[25].

As we observed *BRD9* mis-splicing in a range of cancer types that carry distinct *SF3B1* mutations, targeting *BRD9* mis-splicing could be a productive pan-cancer therapy. Although ASO treatment merely restored *BRD9* mRNA and BRD9 protein to normal levels, we nonetheless observed a strong suppression of tumour growth. The functional effect of correcting *BRD9* mis-splicing was particularly notable given that the UVM models used here contain hundreds of other mis-splicing events and multiple pro-tumorigenic mutations. Given recent clinical successes with treating spinal muscular atrophy and other diseases with ASOs[26], the tumour-suppressive effects of correcting *BRD9* mis-splicing suggest that oligonucleotide-based therapy may prove similarly promising for treating cancers with spliceosomal mutations.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1646-9.

Received: 12 January 2019; Accepted: 30 August 2019;
Published online 9 October 2019.

1. Yoshida, K. et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
2. Papaemmanuil, E. et al. Somatic *SF3B1* mutation in myelodysplasia with ring sideroblasts. *N. Engl. J. Med.* **365**, 1384–1395 (2011).
3. Wang, L. et al. *SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–2506 (2011).
4. Seiler, M., et al. Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell Rep.* **23**, 282–296.e284 (2018).
5. Alpsoy, A. & Dykhuizen, E. C. Glioma tumor suppressor candidate region gene 1 (*GLTSCR1*) and its paralog GLTSCR1-like form SWI/SNF chromatin remodeling subcomplexes. *J. Biol. Chem.* **293**, 3892–3903 (2018).
6. Michel, B. C. et al. A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nat. Cell Biol.* **20**, 1410–1420 (2018).
7. Gatchalian, J. et al. A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. *Nat. Commun.* **9**, 5139 (2018).
8. Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2011).
9. Furney, S. J. et al. *SF3B1* mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* **3**, 1122–1129 (2013).
10. Alsafadi, S. et al. Cancer-associated *SF3B1* mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7**, 10615 (2016).
11. Harbour, J. W. et al. Recurrent mutations at codon 625 of the splicing factor *SF3B1* in uveal melanoma. *Nat. Genet.* **45**, 133–135 (2013).
12. Martin, M. et al. Exome sequencing identifies recurrent somatic mutations in *EIF1AX* and *SF3B1* in uveal melanoma with disomy 3. *Nat. Genet.* **45**, 933–936 (2013).
13. Cretu, C. et al. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. *Mol. Cell* **64**, 307–319 (2016).
14. DeBoever, C. et al. Transcriptome sequencing reveals potential mechanism of cryptic 3′ splice site selection in *SF3B1*-mutated cancers. *PLoS Comput. Biol.* **11**, e1004105 (2015).
15. Darman, R. B. et al. Cancer-associated SF3B1 hotspot mutations induce cryptic 3′ splice site selection through use of a different branch point. *Cell Rep.* **13**, 1033–1045 (2015).
16. Hohmann, A. F. et al. Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat. Chem. Biol.* **12**, 672–679 (2016).
17. Gozani, O., Potashkin, J. & Reed, R. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell. Biol.* **18**, 4752–4760 (1998).
18. Pineda, J. M. B. & Bradley, R. K. Most human introns are recognized via multiple and tissue-specific branchpoints. *Genes Dev.* **32**, 577–591 (2018).
19. Remillard, D. et al. Degradation of the BAF complex factor BRD9 by heterobifunctional ligands. *Angew. Chem. Int. Edn Engl.* **56**, 5738–5743 (2017).
20. Martens, J. A. & Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.* **13**, 136–142 (2003).
21. Brien, G. L. et al. Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. *eLife* **7**, e41305 (2018).
22. Moore, A. R. et al. Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. *Nat. Genet.* **48**, 675–680 (2016).
23. Baldi, A. et al. The HtrA1 serine protease is down-regulated during human melanoma progression and represses growth of metastatic melanoma cells. *Oncogene* **21**, 6684–6688 (2002).
24. Chien, J. et al. Serine protease HtrA1 associates with microtubules and inhibits cell migration. *Mol. Cell. Biol.* **29**, 4177–4187 (2009).
25. Walker, C. J. et al. Genome-wide association study identifies an acute myeloid leukemia susceptibility locus near *BICRA*. *Leukemia* **33**, 771–775 (2019).
26. Stein, C. A. & Castanotto, D. FDA-approved oligonucleotide therapies in 2017. *Mol. Ther.* **25**, 1069–1075 (2017).
27. Pellagatti, A. et al. Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. *Blood* **132**, 1225–1240 (2018).
28. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0, http://www.repeatmasker.org (2013–2015).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

Sample sizes for xenograft experiments were chosen on the basis of published studies of known oncogenic drivers of relevant models (for example, expression of the oncoprotein CYSLTR2(L129Q) in Melan-a cells). Mice were randomly assigned to experimental groups. The data presented did not require the use of blinding.

**Cell lines and tissue culture.** All cell lines underwent short-tandem repeat testing (ATCC) and Memorial Sloan Kettering integrated mutation profiling of actionable cancer targets (MSK IMPACT) genetic analysis[29] to evaluate for spliceosome-gene mutational status and status of recurrently mutated genes in cancer. HEK293T cells were grown in DMEM with 10% FCS. Ba/F3 cells and Melan-a cells (provided by D. Bennett) were grown in RPMI with 10% FCS with 1 ng/ml IL-3 (PeproTech, 213-13) and 200 nM TPA (Sigma-Aldrich), respectively, unless noted otherwise. The K562 and NALM-6 isogenic cell lines (engineered to express *SF3B1*$^{K700E}$, *SF3B1*$^{K666N}$ or *SF3B1*$^{K700K}$ (wild-type control for genome engineering) from the endogenous *SF3B1* locus) were cultured in RPMI with 10% FCS and their generation has previously been described[30]. MEL270, MEL285 and RN2 cell lines were cultured in RPMI with 10% FCS. MEL202, 92-1 and SK-MEL30 cells were grown in RPMI with 10% FCS and 1% GlutaMAX (Gibco). UPMD1 and UPMD2 cells were grown in Ham F-12 with 10% FCS. CFPAC1 cells were cultured in IMDM with 10% FCS. KPC, Miapaca2 and B16 cells were cultured in DMEM with 10% FCS. Panc 05.04 cells were grown in RPMI with 20% FCS and 20 units per millilitre human recombinant insulin. T47D cells were cultured in RPMI1640 supplemented with 10% fetal bovine serum (Corning), 100 μg/ml penicillin, 100 mg/ml streptomycin (Corning), and 4 mM glutamine. All of the cell culture media included penicillin (100 U/ml) and streptomycin (100 μg/ml).

**Primary human samples and human patient-derived xenograft models.** Studies were approved by the Institutional Review Boards (IRBs) of Memorial Sloan Kettering Cancer Center (MSK), informed consent was obtained from all subjects (under MSK IRB protocol 06-107) and studies were conducted in accordance to the Declaration of Helsinki protocol. Patients provided samples after their informed consent, and samples of primary human de-identified chronic lymphocytic leukaemia derived from whole peripheral blood or bone marrow mononuclear cells were used. Patient-derived xenograft models were performed using tumour biopsies from de-identified patients under MSK IRB protocol 14-191. Genomic alterations in melanoma tumour biopsies and chronic lymphocytic leukaemia cells were analysed using the MSK IMPACT[29] assay or FoundationOne Heme[31] assay, both as previously described. Patient samples were anonymized by the Hematologic Oncology Tissue Bank of MSK (for chronic lymphocytic leukaemia samples) and the MSK Antitumour Assessment Core Facility (for patient-derived xenograft samples).

**Mice.** All mice were housed at Memorial Sloan Kettering Cancer Center (MSKCC). All mouse procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSKCC. All mouse experiments were performed in accordance with a protocol approved by the MSKCC Institutional Animal Care and Use Committee (11-12-029). SCID mice (Jackson Laboratories stock no. 001303) were used for all human cell line xenografts and NSG mice (Jackson Laboratories stock no. 005557) were used for patient-derived xenografts. For all mouse experiments, the mice were monitored closely for signs of disease or morbidity daily and were killed when they showed a volume of the visible tumour formation above 1 cm$^3$, failed to thrive, experienced weight loss >10% total body weight or showed open skin lesions, bleeding or any signs of infection. These limits were not exceeded in any experiments.

**HA-tag knock-in into endogenous *BRD9*.** The following guide RNA sequence targeting the *BRD9* transcriptional start site was selected using the optimized CRISPR design tool (http://crispr.mit.edu): CGAGTGGCGCTCGTCCTACG. DNA oligonucleotides were purchased from IDT and cloned into the px458-GFP vector. For homologous recombination, we purchased a custom IDT Ultramer 197-bp repair template (single-stranded donor DNA) with the HA sequence (TACCCATACGATGTTCCAGATTACGCT) directly following the *BRD9* start codon. This 197-bp fragment contained two silent mutations, one to remove the PAM site (AGG > AAG) and another to introduce an XhoI restriction enzyme (CTCGGG > CTCGAG) site upstream of the HA tag. The 197-bp fragment also contained 83 bp of homology to the *BRD9* 5′ UTR upstream of the HA tag and 87 bp of homology to the *BRD9* exon 1 downstream of the start codon. Five micrograms of the targeting construct and 500 nM of the repair template were nucleofected into K562 *SF3B1*$^{K700E}$ cells and MEL270 cells using the Lonza Nucleofector V kit and Program T-003 on the nucleofector device. Nucleofected cells were single-cell-sorted on the basis of GFP positivity 48 h after nucleofection. Clones were screened for the presence of successful HA insertion by *BRD9* exon 1 PCR and subsequent restriction enzyme digestion with XhoI and direct Sanger sequencing. A single positive clone containing the HA coding sequencing was selected to carry out further studies. The sgRNAs were CACCCGAGTGGCGCTCGTCCTACG (top) and AAACCGTAGGACGAGCGCCACTCG (bottom); the single-strand donor DNA was CCAGGGGGCGGTGGCGGCCAAGGTCCGACCGGGTGCCAGCTGTTCCCAGCCCCCGCCTCGAGCCCGCCGCCGGCGCCGCCATGTACCCATACGATGTTCCAGATTACGCTGGCAAGAAGCACAAGAAGCACAAGGCCGAGTGGCGCTCGTCCTACGAAGGTGAGGCGGCGGCGCTTTGTGACGCGCGGCGGGCGGGG; the PCR primers were forward (fwd) AGCGAGCTCGGCAACCTCG and reverse (rev) CTTCAGGACTAGCTTTAGAGGC; the Sanger sequence primer was rev TGCAGCCTCGAACCCAGAAC

**Overexpression of *SF3B1* cDNA in K562 cells.** Two micrograms of PiggyBac Transposase construct (CMV-PB-Transposase-IRES-TK-HSV) and 6 μg of wild-type *SF3B1* (ITR-CAG-Flag-SF3B1$^{WT}$-IRES-Puro-ITR) or *SF3B1*$^{K700E}$ (ITR-CAG-Flag-SF3B1$^{K700E}$-IRES-Puro-ITR) cDNA constructs were electroporated into 2 × 10$^6$ cells (in 200 μl volume) using the Amaxa Nucleofector Protocol (Program T-003) according to manufacturer instructions (Lonza). Puromycin selection (1 μg/ml) was initiated 4 days after electroporation to select for cells that successfully incorporated the constructs. Sanger sequencing was performed to confirm successful integration of the cDNA plasmid using the following primers: fwd, TCCAATCAAAGATCTTCTTCCAA and rev, GAGCAGGTTTCTGCAACGAT.

**RT–PCR and quantitative RT–PCR.** Total RNA was isolated using RNeasy Mini or Micro kit (Qiagen). For cDNA synthesis, total RNA was reverse-transcribed to cDNA with SuperScript VILO cDNA synthesis kit (Life Technologies). The resulting cDNA was diluted 10–20 fold before use. Quantitative RT–PCR (qRT–PCR) was performed in 10-μl reactions with SYBR Green PCR Master Mix. All qRT–PCR analysis was performed on an Applied Biosystems QuantStudio 6 Flex Cycler (ThermoFisher Scientific). Relative gene expression levels were calculated using the comparative $C_T$ method. Primers used in RT–PCR reactions were: *BRD9* (human) fwd, GCAATGACATACAATAGGCCAGA and rev, GAGCTGCCTGTTTGCTCATCA; *Brd9* (mouse) fwd, TTGGAGATGGAAGTCTGCTCT and rev, GCAACTTGCTAGACAGTGAACT; *BRD9* poison exon (human) fwd, AGCTCTGTTCTGGAGTTCATG and rev, CTGAAGAAACTCATAGGGGTCGTG; *Brd9* poison exon (mouse) fwd, GGCCCTGTTCTGGACTTCATG and rev, CTGAAGGAATTCATAAGGGTCGTG; *BRD9* poison exon inclusion for small interfering RNA (siRNA) experiment (human) fwd, CAGCAGCTCTGTTCTGGAGT and rev: CCTGAAAGAAACCAGAGAGCTG; *BRD9* poison exon exclusion for siRNA experiment (human) fwd, CAGCAGCTCTGTTCTGGAGT and rev, TCACCTTCCCCAGAGAGCTG; *EPB49* (also known as *DMTN*) cassette exon inclusion (human) fwd, GCCTGCAGAACGGAGAGG and rev: ACCACTAGCATTTCATAGGGATAGATCT; *EPB49* cassette exon exclusion (human) fwd, GCCTGCAGATCTATCCCTATGAAAT and rev, CTCAAGCCGCATCCGATCC; *BRD9* poison exon for mRNA half-life experiment (human) fwd, GTTGGGGACACCCTAGGAGA, rev (exclusion-specific), CTTCACCTTCCCCAGAGAGC and rev (inclusion-specific), CCCTGAAAGAAACCAGAGAGC; 18S rRNA (human) fwd CTACCACATCCAAGGAAGCA and rev, TTTTTCGTCACTACCTCCCCG; *Mitf* (mouse) fwd, CCAACAGCCCTATGGCTATGC and rev: CTGGGCACTCACTCTCTGC; *Dct* (mouse) fwd, GTCCTCCACTCTTTTACAGACG and rev, ATTCGGTTGTGACCAATGGGT; *Pmel* (mouse) fwd, GAGCTTCCTTCCCGTGCTT and rev, TGCCTGTTCCAGGTTTTAGTTAC; *Tyrp1* (mouse) fwd, CCCCTAGCCTATATCTCCCTTTT and rev, TACCATCGTGGGGATAATGGC; *GAPDH* (human) fwd, GGAGCGAGATCCCTCCAAAAT and rev, GGCTGTTGTCATACTTCTCATGG; and *Gapdh* (mouse) fwd, AGGTCGGTGTGAACGGATTTG and rev, TGTAGACCATGTAGTTGAGGTCA.

**mRNA stability assay.** For mRNA half-life measurement using qRT–PCR, K562 and NALM-6 cells with isogenic *SF3B1*$^{K700E}$ mutations were infected with anti-*UPF1* shRNAs or control shRNA, and treated with 2.5 μg/ml actinomycin D (Life Technologies) and collected at 0, 2, 4, 6 and 8 h (using methods as previously described[32]). *BRD9* poison exon inclusion or exclusion and 18S rRNA mRNA levels were measured by qRT–PCR.

**Western blotting.** For western blotting, the following antibodies to the following proteins were used: BRD9 (Bethyl Laboratories A303-781A and Active Motif 61538), SF3B1/Sap-155 (MBL D221-3), Flag-M2 (Sigma-Aldrich F-1084), β-actin (Sigma-Aldrich A-5441), GLTSCR1 (Santa Cruz Biotechnology sc-515086), GLTSCR1L (Thermo Fisher Scientific PA5-56126), BRM (Bethyl Laboratories A303-015A), BRG1 (Santa Cruz Biotechnology sc-17796), BAF155 (Santa Cruz Biotechnology sc-48350), BAF60A (Santa Cruz Biotechnology sc-135843), BAF47 (Santa Cruz Biotechnology sc-166165), ARID1A (Santa Cruz Biotechnology sc-373784), ARID2 (Santa Cruz Biotechnology sc-166117), BRD7 (Thermo Fisher Scientific PA5-49379), U2AF2 (Bethyl Laboratories A303-665A), U2AF1 (Bethyl Laboratories A302-080A), histone H3 (Abcam ab1791), HTRA1 (R&D Systems MAB2916-SP). All primary antibodies for western blotting were diluted to a final concentration of 1:500 to 1,000, in either 5% BSA (Sigma-Aldrich) in 0.05% TBS–Tween 20 (TBS-T) or 5% skim milk in 0.05% TBS-T. Nuclear extracts were quantified using BCA and 1 mg protein (1 mg ml$^{-1}$ in immunoprecipitation buffer supplemented with protease inhibitors) was used per immunoprecipitation.

Proteins were incubated for 3 h with 2–5 μg of antibody or with protein A/G PLUS-Agarose (Santa Cruz Biotechnology sc-2003) with rotation at 4 °C. After washing three times with Pierce IP Lysis Buffer (Thermo Fisher Scientific 87787), immunoprecipitated proteins were eluted with Pierce Lane Marker Reducing Sample Buffer (Thermo Fisher Scientific 39000) and loaded onto 4–12% Bis-Tris NuPAGE Gels (Life Technologies).

**ChIP.** For ChIP–seq studies in MEL270 cells, antibodies to endogenous BRG1 (Abcam EPNCIR111A, lot no. GR3208604-8), GLTSCR1 (Santa Cruz SC-240516, lot no. A2313) and BRD9 (Abcam ab137245) were used, and ChIP was performed as previously described in detail[6]. MEL270 cells transduced with empty vector, doxycycline-inducible wild-type *SF3B1* cDNA, or doxycycline-inducible *SF3B1*$^{K700E}$ cDNA in the backbone of pInducer20, were treated with doxycycline (1 μg/ml) plus BRD9 degrader[19] (250 nM) or DMSO for 72 h before crosslinking.

**Mass spectrometry.** For anti-Flag–BRD9 ChIP followed by mass spectrometry, K562 cells transduced with empty vector or 3×Flag-tagged BRD9 were grown in RPMI with 10% FCS. Ten million cells were crosslinked according to the manufacturer's instruction (Active Motif) and as previously described[33]. Cells were fixed with 1% methanol-free formaldehyde (Sigma, F-8775) for 8 min and quenched with 0.125 M glycine (Sigma, G-7403). Chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of 300–500 bp. Genomic DNA (input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for reverse-crosslinking, followed by ethanol precipitation. Pellets were resuspended and the resulting DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantification of the total chromatin yield. An aliquot of chromatin (150 μg) was precleared with protein G agarose beads (Invitrogen). Proteins of interest were immunoprecipitated using 15 μg of antibody against Flag and protein G magnetic beads. Protein complexes were washed and trypsin was used to remove the immunoprecipitate from beads and digest the protein sample. Protein digests were separated from the beads and purified using a C18 spin column (Harvard Apparatus). The peptides were vacuum-dried using a SpeedVac. Digested peptides were analysed by liquid chromatography and tandem mass spectrometry on a Thermo Scientific Q Exactive Orbitrap mass spectrometer in conjunction with a Proxeon Easy-nLC II HPLC (Thermo Scientific) and Proxeon nanospray source.

Protein identifications were accepted if they contained at least one identified peptide. Proteins that contained similar peptides and could not be differentiated on the basis of tandem mass spectrometry analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. Protein peptide evidence is specified in Supplementary Table 9. The final list was generated by taking all proteins with a spectral count of five and above from each replicate reaction and comparing them in a Venn diagram against IgG control replicates. Proteins unique to both experimental replicates were then applied to the PANTHER database for protein ontology results.

**shRNA experiments.** Cells were transduced with a doxycycline-inducible LT3GEPIR lentiviral vector, T3G-GFP-mirE-PGK-Puro-IRES-rtTA3[34], expressing shRNAs against *BRD9* or a non-targeting shRNA against *Renilla*. shRNAs were induced with the addition of doxycycline (2.0 μg/ml, Sigma Aldrich). All shRNAs were designed using the SplashRNA algorithm[35]. The shRNA sequences are: *BRD9* shRNA no. 1 (human, shBRD9_352): TTTATTATCATTGAATATCCAG; *BRD9* shRNA no. 2 (human, shBRD9_353): TTTTATTATCATTGAATATCCA; *Brd9* shRNA no. 1 (mouse, shBrd9_511): TTTATTATCATTGAATACCCAG; *Brd9* shRNA no. 2 (mouse, shBrd9_512): TTTTATTATCATTGAATACCCA; *HTRA1* shRNA no. 1 (human, shHTRA1_1192): TTTTTAATCTTATCAGATGGGA; *HTRA1* shRNA no. 2 (human, shHTRA1_1669): TGAACAAACAAAATGGCAGTCA; and *HTRA1* shRNA no. 3 (human, shHTRA1_1898): TTCTATCTACGCATTGTATCGA.

**siRNA transfections.** K562 cells were transfected with a non-targeting control siRNA (Dharmacon, D-001810-01, target sequence: UGGUUUACAUGUCGACUAA), an siRNA pool against *U2AF1* (Dharmacon ON-TARGETplus SMARTpool, L-012325-01) or an siRNA pool against *U2AF2* (Dharmacon ON-TARGETplus SMARTpool, L-012380-02) using the Nucleofector II device from Lonza with the Cell Line Nucleofector Kit V (program T16). RNA and protein were extracted 48 h after transfection. cDNA was produced using 1 μg of RNA and the Superscript III first strand synthesis system (Thermo Fisher, 18080051).

**In vitro competition assay.** For the competition assay to evaluate the cellular effect of sgRNA against *BRD9* or *Brd9*, cell lines were transduced with LentiCas9-Blast (Addgene no. 52962) and then single-cell-sorted into 96-well plates. Among these clones, we used single clones with strong Cas9 expression, which was confirmed by western blotting. Cas9-expressing cells were lentivirally transduced with iLenti-guide-GFP vectors subcloned with the gRNA sequences, in which sgRNA expression was linked to GFP expression. The percentage of GFP-expressing cells was then measured over time after infection using BD LSRFortessa. The GFP-positive rates in living cells at each point compared to that of day 2 were calculated. Similarly, to evaluate the cellular effect of *BRD9* fragment cDNA, *HTRA1* cDNA or shRNAs against *BRD9* or *HTRA1*, GFP-positive rates were measured after transducing pMIGII-backbone plasmids (cDNA) and LT3GEPIR plasmids (shRNA), respectively, into melanoma cell lines. The sgRNAs were as follows: *BRD9* sgRNA no. 1 (human): ACTCCAGTTACTATGATGAC, *BRD9* sgRNA no. 2 (human): AGAGAGGGAGCACTGTGACA, *BRD9* sgRNA no. 3 (human): AGATACCGTGTACTACAAGT, *Brd9* sgRNA no. 1 (mouse): ATTAACCGGTTTCTCCCGGG, *Brd9* sgRNA no. 2 (mouse): GGAACACT GCGACTCAGAGG, *Brd9* sgRNA no. 3 (mouse): ACTTGCTAGACAGTGAACTC, and control (scrambled sgRNA): ACGGAGGCTAAGCGTCGCAA.

**CRISPR enrichment screening for NMD targets.** First, Ba/F3 cells were transduced with LentiCas9-Blast (Addgene no. 52962) and single-cell-sorted into 96-well plates. Among these clones, we used a single clone with strong Cas9 expression. The sgRNA library of NMD targets in SF3B1-mutant cells were amplified and packaged as lentivirus. The library includes 4 sgRNAs against each target gene (a total of 274 genes), 100 control sgRNAs and positive-control sgRNAs against *Pten*. Ba/F3 cells were transduced with the lentivirus-carrying sgRNA library produced by 293FT cells, and puromycin selection (2 μg/ml) was performed in IL-3-containing medium for 7 days. Then, we washed out IL-3 (on day 0) and the surviving cells were collected 7 days after IL-3 depletion (day 7). Cell pellets were lysed and genomic DNA was extracted (Qiagen), and quantified by Qubit (Thermo Scientific). A quantity of gDNA covering a 1,000× representation of gRNA was PCR-amplified using Q5 high-fidelity polymerase (NEB cat. no. M0491) to add Illumina adapters and multiplexing barcodes. Amplicons were quantified by Qubit and Bioanalyzer (Agilent) and sequenced on Illumina HiSeq 2500. Sequencing reads were aligned to the screened library; counts were computed for each gRNA and counts for each sgRNA were compared between days 0 and 7 after cytokine depletion. For the probe-level analysis, we fitted a negative binomial generalized log-linear model and performed a likelihood ratio test with glmFit and glmLRT in the Bioconductor edgeR package. For the gene-level analysis, we used the CAMERA test as implemented in edgeR[36–38]. FDR values were computed using the Benjamini–Hochberg method.

**CRISPR-directed mutations.** Cas9-expressing MEL202 and MEL270 cells were transduced with iLenti-guide-Puro vector targeting the 5′ end of the *BRD9* poison exon. The sgRNA used to induce mutations was: AAAATACTCAGTTTCTTTCA. Single-cell sorting was performed into a 96-well plate using a BD FACSAria III cell sorter after puromycin selection (2.0 μg/ml for 7 days). The mutations caused by Cas9 and the sgRNA were confirmed by PCR amplification, followed by Sanger sequencing. The PCR primers were fwd, TGTTGGGTCAGGAAGAGACTTG and rev, CCATGGACTGAACGGATTCC; and the Sanger sequencing primer was fwd, TGTTGGGTCAGGAAGAGACTTG.

**Colony-forming assays.** Single-cell suspensions were prepared from MEL202 and MEL270 cells with or without CRISPR-mediated insertions and deletions (indels), 3,000 cells from each cell line were plated in triplicates in 6-well treated plates and colonies were enumerated 10 days later. After 10 days, colonies were fixed with 3.7% paraformaldehyde for 5 min and stained in a solution of 0.05% crystal violet for 30 min at room temperature and washed in PBS and tap water.

**In vitro cell viability assays.** Cells were seeded in white flat-well 96-well plates (Costar) at a density of 1,000 cells per well. ATP luminescence readings were taken every 24 h after seeding, using Cell Titer Glo (Promega) according to the manufacturer's instructions.

***BRD9* minigene assay.** We identified putative exonic splicing enhancers with SpliceAid 2[39]. We used SF3B1 wild-type MEL270 and T47D cells transduced with Flag–SF3B1(WT) or Flag–SF3B1(K700E) (in the backbone of pInducer20 (Addgene no. 44012)). After drug selection with neomycin (Thermo Fisher Scientific 10131027), the selected cells were treated with 1 μg /ml doxycycline (Sigma D9891). The *BRD9* minigene construct was generated by inserting the DNA fragment containing the *BRD9* genomic sequence from exon 14 to exon 15 in between the BamHI and AgeI restriction sites in the FRE5 plasmid (Addgene 62377) via Gibson assembly. *BRD9* minigene mutagenesis was performed with the Agilent QuikChange II site-directed mutagenesis kit with specific primers according to the manufacturer's directions. For transient transfection experiments, cells were seeded into a 24-well plate one day before transfection of *BRD9* minigene constructs in the presence of X-tremeGENE HP DNA transfection reagent (Roche) according to the manufacturer's directions. Forty-eight hours after transfection, cells were collected and RNA was extracted using Qiagen RNeasy mini kit. Minigene-derived and endogenous BRD9 transcripts were analysed by RT–PCR using specific primers. Primers and oligonucleotides used in RT–PCR reactions were: cloning fwd GGACCCAGTACCAGGATCCGTTGGGGACACCCTAGGAG and rev CTTGGAGGAGCGCACACCGGTCAGGTGGTGCTGGTCCCTC; RT–PCR fwd (minigene) GGATTACAAGGATGACGATGAC, fwd (endogenous) CATGAAGCCTCCAGATGAAG and rev (common) CTTCGTCGT CTCATCCAAGTTC; mutagenesis (AAAA to TTTT) fwd CAAAGGGATATA

TTTTGAGATTTTTTACTCAGTTTCTTTCAGG and rev CCTGAAAGAAACTGAGTAAAAAATCTCAAAATATATCCCTTTG; mutagenesis (AAAA to ATAT) fwd CAAAGGGATATATTTTGAGATATATTACTCAGTTTCTTTCAGG and rev CCTGAAAGAAACTGAGTAATATATCTCAAAATATATCCCTTTG; mutagenesis (GATAAAA to TTTTTTT) fwd GTCAAAGGGATATATTTTGATTTTTTTTACTCAGTTTCTTTCAGG and rev ; mutagenesis (TTTCT to TTACA at exon 14a) fwd GAGATAAAATACTCAGTTACATTCAGGGCCTGCCATCTATC and rev GATAGATGGCAGGCCCTGAATGTAACTGAGTATTTTATCTC; mutagenesis (TTTCT to TTGCG at exon 14a) fwd GAGATAAAATACTCAGTTGCGTTCAGGGCCTGCCATCTATC, and rev GATAGATGGCAGGCCCTGAACGCAACTGAGTATTTTATCTC; mutagenesis (TTTCT to TTTCA at exon 14a) fwd GAGATAAAATACTCAGTTTCATTCAGGGCCTGCCATCTATC and rev GATAGATGGCAGGCCCTGAATGAAACTGAGTATTTTATCTC; mutagenesis (TTTCT to TTTCG at exon 14a) fwd GAGATAAAATACTCAGTTTCGTTCAGGGCCTGCCATCTATC, and rev GATAGATGGCAGGCCCTGAACGAAACTGAGTATTTTATCTC; mutagenesis (TTTCT to TTTAT at exon 14a) fwd GAGATAAAATACTCAGTTTATTTCAGGGCCTGCCATCTATC and rev GATAGATGGCAGGCCCTGAAATAAACTGAGTATTTTATCTC; and mutagenesis (branchpoint A to G) fwd GTCAAAGGGATATATTTTGGGATAAAATACTCAGTTTCTTTC and rev GAAAGAAACTGAGTATTTTATCCCAAAATATATCCCTTTGAC.

**Lariat sequencing.** To map the branchpoints that were used when the *BRD9* poison exon was included or excluded, RT–PCR was performed to amplify branchpoint-spanning fragments from lariat RNAs arising during normal (poison exon exclusion) or aberrant (poison exon inclusion) splicing of *BRD9* pre-mRNA. In brief, SuperScript III reverse transcriptase (Invitrogen) and a primer complementary to the intronic sequences downstream of the 5′ splice sites were used to generate cDNA from lariat RNAs. Branchpoint-spanning fragments were then amplified from lariat RNAs by nested PCR with pairs of outer primers (with the RT primer being the reverse primer) and inner primers. The forward primers were complementary to sequences about 200–300 nucleotides upstream of the 3′ splice sites and the reverse primers were complementary to sequences downstream of the 5′ splice sites (Supplementary Table 11). The PCR products were cloned into the pGEM-T vector (Promega) and sequenced by Sanger sequencing.

**Melanoma transplant model.** We resuspended stably transduced Melan-a, MEL270 and MEL202 cells (1M cells) with doxycycline-inducible shRNAs, cDNAs or sgRNAs in 100 μl of a 1:1 mix of medium and Matrigel (BD Biosciences), and subcutaneously and bilaterally injected the mix into the flanks of 7-week-old female CB17-SCID mice (Taconic). For doxycycline-regulated shRNA induction, we used doxycycline-containing diets (625 mg/kg diet, Envigo). To assess tumour growth, at least five mice per group were injected for a total of ten tumours per group. No randomization or blinding was used in the analysis of tumour growth. Tumours were measured with callipers every seven days. Growth curves were visualized with Prism GraphPad 8.0. Tumour volume was calculated using the formula; Volume = π(length)(width)(height)/6.

**In vitro morpholino transfection.** To deliver morpholinos into cultured cell lines, we followed the manufacturer's instruction (GeneTools). In brief, we used 6 μM Endo-Porter after adding morpholinos (final concentration of 10 μM). RNA and proteins are collected 48 h after delivery. Morpholino target sequence no. 3 was TAATGAGGCAAGTCCAGTCCCGCTT; no. 6 was AAAGAGGGGATAATGAGGCAAGTCC; and no. 7 was GGGATAATGAGGCAAGTCCAGTCCC.

**In vivo morpholino treatment.** Treatment with morpholinos was started when the tumour volume in mice reached 100–200 mm$^3$. Cohorts were treated intratumorally with 12.5 mg/kg scrambled or poison-exon-targeting Vivo-Morpholinos (AAAGAGGGGATAATGAGGCAAGTCC, GeneTools) dissolved in 50 μl PBS, every 2 days for 8 doses in total. The mice were dissected 24 h after the final treatment. For the patient-derived xenograft model, patient-derived rectal melanoma cells (*SF3B1*$^{R625C}$) and UVM cells (*SF3B1* wild type) were serially transplanted into SCID mice and treated similarly.

**In vivo metastasis model.** For lung experimental metastasis, B16 and 92.1 melanoma cells retrovirally transduced with shRNAs targeting *Renilla*, *BRD9* (no. 1 and no. 1) or *Brd9* (no. 1 or no. 2) in MLS-E vector (sorted using GFP) were trypsinized, resuspended in PBS and then 0.4 M cells in 0.2 ml PBS were injected via the lateral tail vein using a 27-gauge needle. Mice were killed 14 days after injection and tissues were isolated and fixed in 4% paraformaldehyde (Thermo Fisher Scientific). For evaluation of metastatic colonization of the lung using 92.1 human UVM cells, the burden of metastatic cells was evaluated using GFP expression by flow cytometry as well as anti-GFP immunohistochemistry 14 days following tail-vein injection of 0.4 M cells into NOD–SCID *Il2rg*$^{-/-}$ mice.

**Histological analysis.** Tissues were fixed in 4% paraformaldehyde, processed routinely in alcohol and xylene, embedded in paraffin, sectioned at 5-μm thickness and stained with H&E. Immunohistochemistry was performed on a Leica Bond RX automated stainer (Leica Biosystems). Following heat-induced epitope retrieval at pH 6.0, the primary antibody against Ki67 (Vector VP-K451) was applied, followed by application of a polymer detection system (DS9800, Novocastra Bond Polymer Refine Detection, Leica Biosystems) in which the chromogen was 3,3-diaminobenzidine tetrachloride (DAB) and the counterstain was haematoxylin. Photomicrograph examination of all H&E and immunohistochemistry slides were performed using a Zeiss Axioskop imaging.

***BRD9* expression correlates.** The cor.test (in R) was used to calculate Spearman's $\rho$ and the $P$ value associated with the correlation of *BRD9* expression with the expression of each coding gene across all samples within each cohort from the TCGA. Analysis was restricted to coding genes that are not on the same chromosome arm as *BRD9* (chromosome 5p) to remove potential confounding effects of local correlations. Coding genes with $P < 0.01$ in at least 10 cancer types were ranked by their absolute mean value of $\rho$ (computed across all TCGA cohorts) and classified as RNA-binding if they were annotated with the 'RNA-binding' Gene Ontology term (GO: 0003723).

***BRD9* alternative splicing.** Potential NMD-targeted isoforms of *BRD9* were identified as follows: we queried the MISO v.2.0 alternative splicing annotation[40] for exon skipping and competing splice site events within the *BRD9* gene locus, restricted to those events with evidence of alternative splicing based on spliced junction reads (described in 'Genome annotation, RNA-seq read mapping, and estimation of gene and isoform expression'), assigned open reading frames for the isoforms resulting from each alternative splicing event based on the *BRD9* isoform with the longest open reading frame, and classified isoforms as predicted NMD substrates if they contained a termination codon >50 nt upstream of a splice junction.

Robust linear modelling of *BRD9* expression on the basis of the identified alternatively spliced isoforms of *BRD9* that are predicted NMD substrates was performed for each TCGA cohort with the rlm function in the MASS package in R. Relative expression of each isoform in each sample was estimated from RNA-seq data across all TCGA cohorts as described in 'Genome annotation, RNA-seq read mapping, and estimation of gene and isoform expression'. A $z$-score normalization was performed across all samples for each isoform in each cohort before model fitting. The resulting coefficients from the fitted models were subsequently used to predict *BRD9* expression from *BRD9* NMD-targeted isoform expression.

**RNA-seq library preparation.** RNA-seq libraries were prepared from TRIzol-isolated (Thermo Fisher cat. no. 15596026) RNA using the Illumina TruSeq RNA Library Prep Kit v.2 (Illumina cat. no. RS-122-2001/2). K562 libraries were sequenced at MSKCC with 101-bp single-end reads. MEL270 libraries were sequenced by the FHCRC Genomics Shared Resource with 2 × 51-bp paired-end reads.

**ChIP–seq library preparation.** ChIP–seq libraries were prepared and sequenced as previously described[6] by the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute with 75-bp single-end reads.

**Genomic analysis of SWI–SNF complex members from TCGA.** Mutational analysis of genes encoding members of the SWI–SNF complex was performed as previously described[41].

**Genome annotation, RNA-seq read mapping, and estimation of gene and isoform expression.** RNA-seq reads were processed for gene expression and isoform ratio quantification as previously described[42]. In brief, RNA-seq reads were aligned to the hg19/GRCh37 assembly of the human genome using a gene annotation created by merging the UCSC knownGene gene annotation[43], Ensembl v.71.1 gene annotation[44] and MISO v.2.0 isoform annotation[40]. Read alignment and expression estimation were performed with RSEM v.1.2.4[45], Bowtie v.1.0.0[46] and TopHat v.2.1.1[47]. Isoform ratios were quantified with MISO v.2.0[40]. Gene expression estimates were normalized by applying the trimmed mean of $M$ values method[48] to coding genes. Statistical tests for differential gene and isoform expression were performed for single-sample comparisons with Wagenmakers' Bayesian framework[49] and for sample group comparisons with the Mann–Whitney $U$-test. RNA-seq read-coverage plots (for example, Fig. 1e) represent reads normalized by the number of reads mapping to all coding genes in each sample (per million).

**RNA-seq coverage plots.** RNA-seq coverage plots were made using the UCSC Genome Browser[50] and/or the ggplot2 package in R[51]. Repetitive elements were annotated by RepeatMasker[28].

**Cluster analysis.** Unsupervised clustering of chronic lymphocytic leukaemia, myelodysplastic syndrome and UVM samples (Fig. 1a) was based on the 40 events that were differentially spliced in isogenic UVM (MEL270 cells) as well as myeloid leukaemia (K562 cells) cells expressing *SF3B1*$^{K700E}$ versus wild-type *SF3B1*, restricted to the 30 of these events that had sufficient read coverage in all cohorts for clustering.

**ChIP–seq data analysis.** ChIP–seq reads were mapped to the genome by calling Bowtie v.1.0.0[46] with the arguments '-v 2 -k 1 -m 1–best–strata'. Peaks were called using MACS2 v.2.1.1.20160309[52] against input control libraries with $P < 10^{-5}$ and subsequently filtered to remove peaks contained within ENCODE blacklisted regions[53] and the mitochondrial genome. Subsequent data analysis was performed with Bioconductor in the R programming environment[54]. Consensus peaks between samples were called using the soGGI package v.1.14.0[55]. Peaks were annotated using the ChIPseeker package v.1.18.0[56]. Potential transcription factor binding in a 300-nucleotide region around the centre of consensus peaks was scored using the TFBSTools package v.1.20.0[57], with models taken from the HOCOMOCO v.11 human core collection[58] and applied with a threshold of $P < 10^{-4}$. The highest scores for each consensus peak region were collated for each transcription factor. A two-sided Mann–Whitney $U$-test was used to assess the significance of the difference in scores between constitutive and sensitive peaks for each transcription factor.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

RNA-seq and ChIP–seq data generated as part of this study were deposited in the Gene Expression Omnibus (accession number GSE124720). RNA-seq data from published studies were downloaded from CGHub (TCGA UVM[59]), EMBL-EBI ArrayExpress (Illumina Human BodyMap 2.0: E-MTAB-513), the Gene Expression Omnibus (accession numbers GSE72790 and GSE114922 for chronic lymphocytic leukaemia[15] and myelodysplastic syndromes[27], respectively), or directly obtained from the authors (for UVM[10]). Gel source data can be found in Supplementary Fig. 1. Other data that support the findings of this study are available from the corresponding authors upon reasonable request.

29. Cheng, D. T. et al. Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn*. **17**, 251–264 (2015).
30. Seiler, M. et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. *Nat. Med*. **24**, 497–504 (2018).
31. Leeksma, A. C. et al. Clonal diversity predicts adverse outcome in chronic lymphocytic leukemia. *Leukemia* **33**, 390–402 (2019).
32. Kim, E. et al. *SRSF2* mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* **27**, 617–630 (2015).
33. Mohammed, H. et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep*. **3**, 342–349 (2013).
34. Fellmann, C. et al. An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep*. **5**, 1704–1713 (2013).
35. Pelossof, R. et al. Prediction of potent shRNAs with a sequential classification algorithm. *Nat. Biotechnol*. **35**, 350–353 (2017).
36. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
37. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. *Nucleic Acids Res*. **40**, 4288–4297 (2012).
38. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. **40**, e133 (2012).
39. Piva, F., Giulietti, M., Burini, A. B. & Principato, G. SpliceAid 2: a database of human splicing factors expression data and RNA target motifs. *Hum. Mutat*. **33**, 81–85 (2012).
40. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
41. Mashtalir, N., et al. Modular organization and assembly of SWI/SNF family chromatin remodeling complexes. *Cell* **175**, 1272–1288.e1220 (2018).
42. Dvinge, H. et al. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. *Proc. Natl Acad. Sci. USA* **111**, 16802–16807 (2014).
43. Meyer, L. R. et al. The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res*. **41**, D64–D69 (2013).
44. Flicek, P. et al. Ensembl 2013. *Nucleic Acids Res*. **41**, D48–D55 (2013).
45. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
46. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. **10**, R25 (2009).
47. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
48. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. **11**, R25 (2010).
49. Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H. & Grasman, R. Bayesian hypothesis testing for psychologists: a tutorial on the Savage-Dickey method. *Cognit. Psychol*. **60**, 158–189 (2010).
50. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res*. **12**, 996–1006 (2002).
51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2016).
52. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol*. **9**, R137 (2008).
53. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
54. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
55. Dharmalingam G. & Carroll, T. soGGi: Visualise ChIP–seq, MNase-seq and motif occurrence as aggregate plots summarised over grouped genomic intervals. R package version 1.14.0, https://rdrr.io/bioc/soGGi/ (2018).
56. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
57. Tan, G. & Lenhard, B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).
58. Kulakovskiy, I. V. et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res*. **46**, D252–D259 (2018).
59. Robertson, A.G., et al. Integrative analysis identifies four molecular and clinical subsets in uveal melanoma. *Cancer cell* **32**, 204–220.e215 (2017).
60. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. **20**, 110–121 (2010).
61. McLean, C. Y. et al. GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol*. **28**, 495–501 (2010).

**Acknowledgements** We are grateful for the support of the MSK RNAi core facility for help with the CRISPR screens performed in the study, the MSK anti-tumor assessment core facility for help with patient-derived xenograft experiments and the Genomics Shared Resource of the Fred Hutchinson/University of Washington Cancer Consortium (P30 CA015704). D.I., S.C.-W.L., A.Y. and O.A.-W. are supported by the Leukemia & Lymphoma Society. D.I. is supported by grants from Lydia O'Leary Memorial Pias Dermatological Foundation and Kobayashi Foundation for Cancer Research. A.Y. is supported by grants from the Aplastic Anemia and MDS International Foundation (AA&MDSIF) and the Lauri Strauss Leukemia Foundation. S.X.L. is supported by a Conquer Cancer Foundation and ASCO Young Investigator Award, an Aplastic Anemia & Myelodysplastic Syndrome International Foundation research award, as well as an AACR Lymphoma Research Fellowship. G.-L.C. is a Mahan Fellow. O.A.-W. is supported by the Pershing Square Sohn Cancer Research Alliance, the Henry & Marilyn Taub Foundation and the Starr Cancer Consortium. R.K.B. is a Scholar of The Leukemia and Lymphoma Society (1344-18) and is supported in part by the US National Institutes of Health (R01 DK103854). O.A.-W. and R.K.B. are supported by the Evans MDS Foundation, the US National Institutes of Health (R01 HL128239) and the Department of Defense Bone Marrow Failure Research Program (BM150092 and W81XWH-12-1-0041). The results shown here are in part based upon data generated by the TCGA research network (https://cancergenome.nih.gov/).

**Author contributions** D.I., G.-L.C., B.L., B.C.M., C.K., O.A.-W. and R.K.B. designed the study. D.I., B.L. and J.P. performed minigene assays. D.I., B.C.M. and C.K. performed ChIP–seq experiments. A.R.D., G.-L.C. and C.K. performed ChIP–seq analyses. G.-L.C., K.N., A.P. and R.K.B. performed computational analyses of the CRISPR screen, RNA splicing and gene expression data. D.I. and S.X.L. performed the CRISPR screen. D.I. and B.L. generated and validated the anti-BRD9 PE morpholinos in in vitro assays. D.I., L.B., A.B., S.C.-W.L., A.Y. and H.C. performed the mouse experiments. A.R.M., D.I. and L.E.-H. performed experiments using pancreatic cancer cell lines. T.H. and Y.C. provided melanoma cell lines. D.I., G.-L.C., O.A.-W. and R.K.B. prepared the manuscript with help from all co-authors. J.T. and O.A.-W. provided clinically annotated samples from the MSK Hematology/Oncology Tissue Bank and from the MSK Antitumor Assessment Core Facility. O.A.-W. and R.K.B. provided funding and study supervision.

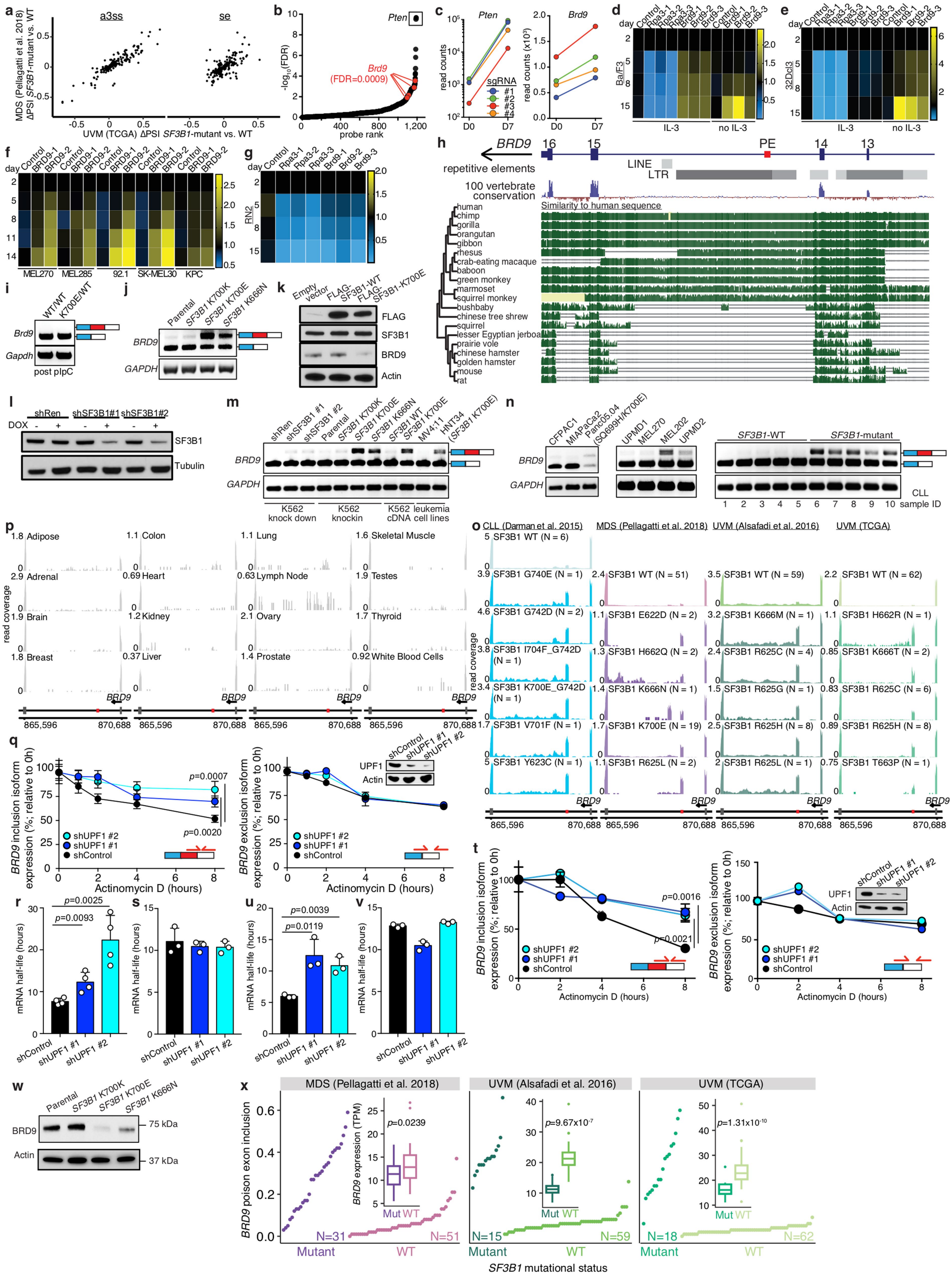
**Competing interests** C.K. is a Scientific Founder, fiduciary Board of Directors member, Scientific Advisory Board member, consultant and shareholder of Foghorn Therapeutics, none of which are related to the current manuscript. O.A.-W. has served as a consultant for H3 Biomedicine, Foundation Medicine, Merck and Janssen; O.A.-W. has received personal speaking fees from Daiichi Sankyo. O.A.-W. has received previous research funding from H3 Biomedicine unrelated to the current manuscript. D.I., O.A.-W. and R.K.B. are inventors on a provisional patent application submitted by the Fred Hutchinson Cancer Research Center that covers BRD9 activation in cancer.

### Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-019-1646-9.

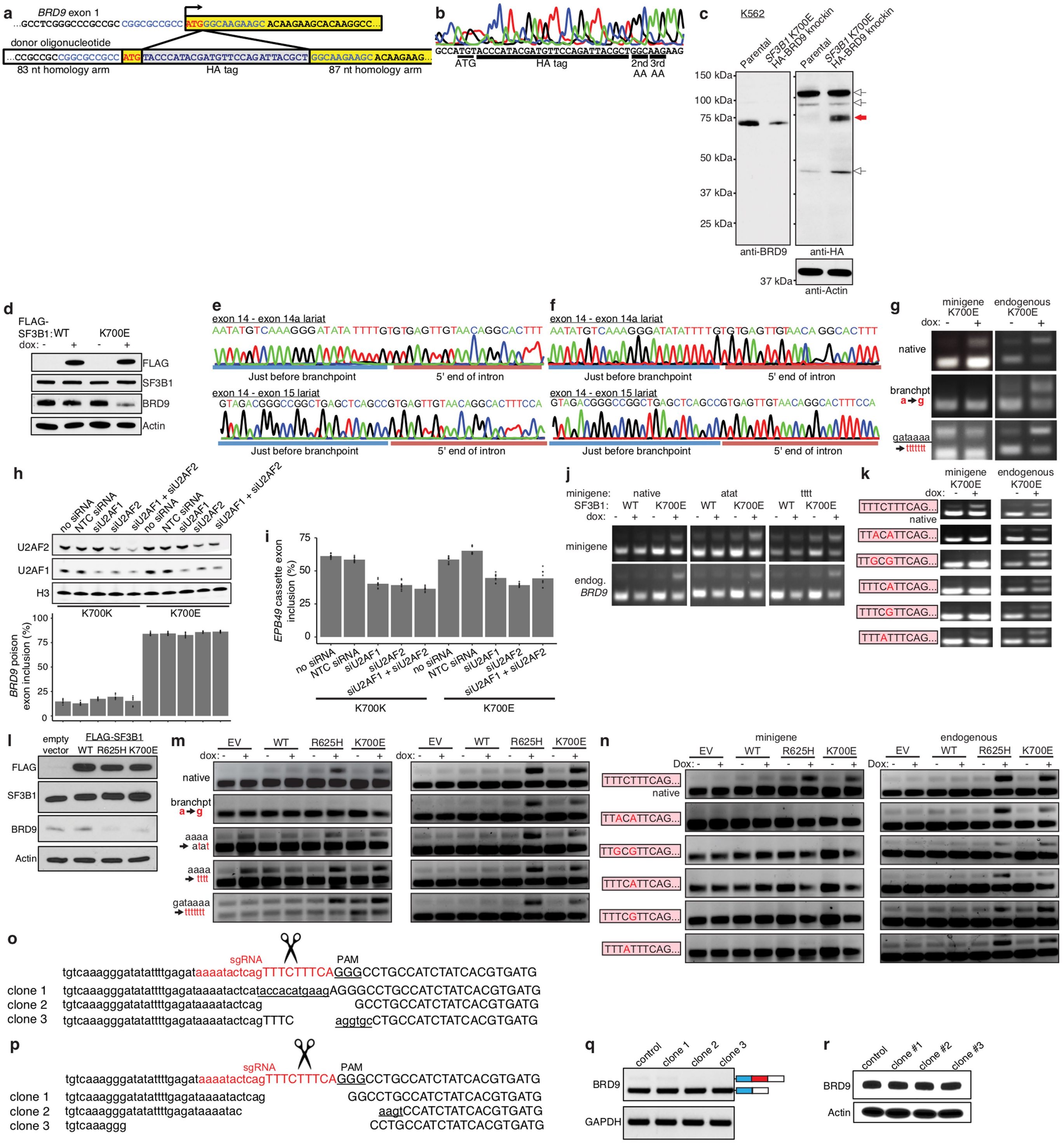
**Correspondence and requests for materials** should be addressed to O.A. or R.K.B.

**Peer review information** *Nature* thanks Boris Bastian, Rotem Karni and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



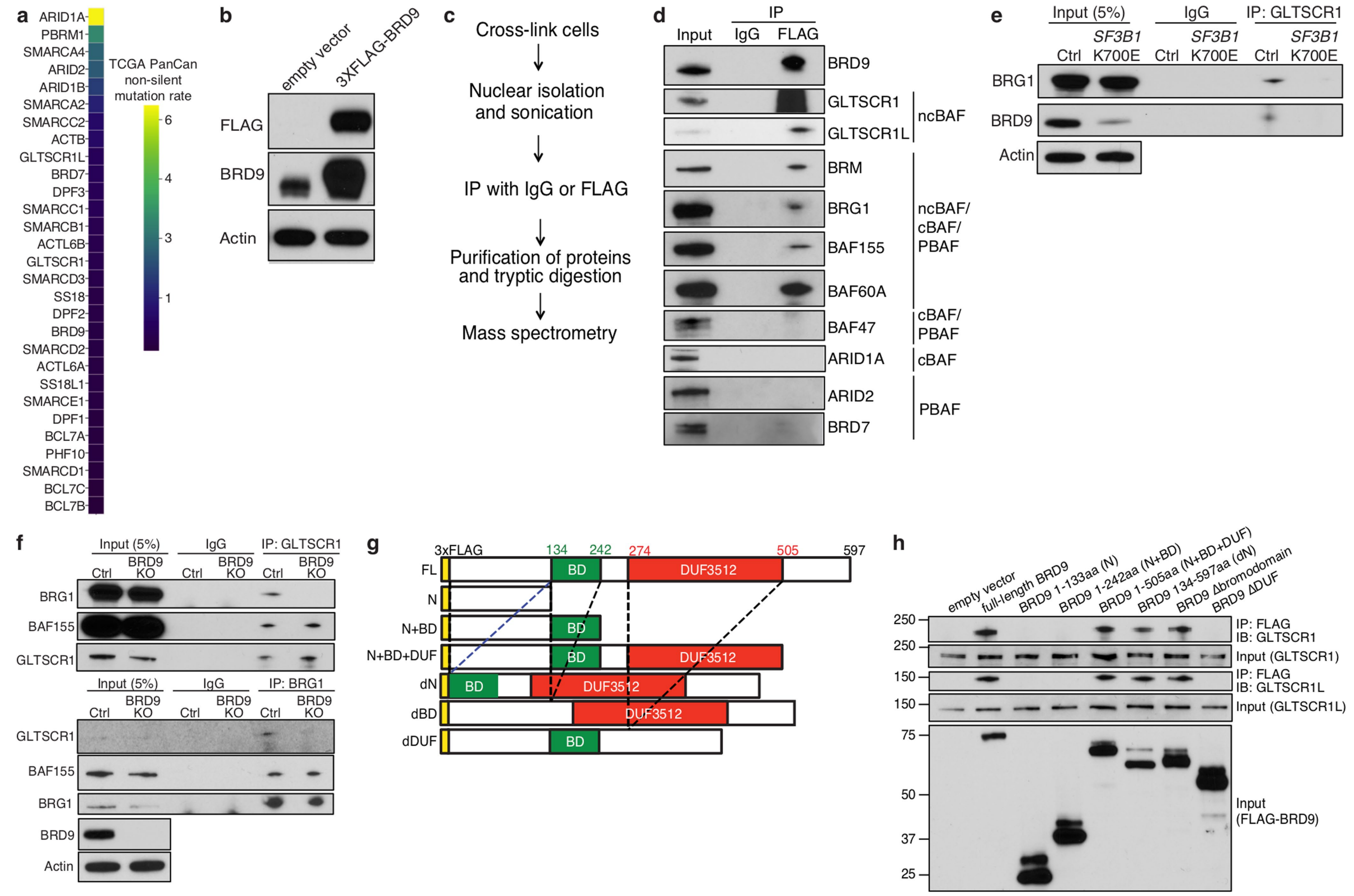
**Extended Data Fig. 1 |** See next page for caption.

**Extended Data Fig. 1 | *BRD9* is mis-spliced in *SF3B1*-mutated human cells, and BRD9 loss confers a proliferative advantage. a**, Scatter plots comparing differential splicing ($\Delta$PSI) between patients in the TCGA UVM cohort with *SF3B1* mutations or wild-type *SF3B1*[59] ($x$ axis) and patients from a myelodysplastic syndromes cohort with *SF3B1* mutations or wild-type *SF3B1*[27] ($y$ axis). Events were classified as alternative 3′ splice sites or skipped exons. **b**, Rank plot for the $-\log_{10}$(FDR) associated with each sgRNA in our CRISPR–Cas9 positive-selection screen. sgRNAs targeting the positive control (*Pten*) and *Brd9* are highlighted. For the probe-level (per-sgRNA) analysis, we fitted a negative binomial generalized log-linear model and performed a likelihood ratio test. FDR values were computed using the Benjamini–Hochberg method. **c**, Read counts for sgRNAs targeting the positive control (*Pten*) or *Brd9*. D0 and D7 indicate days following withdrawal of IL-3. **d**, Heat map summarizing the results of a competition assay to measure the effect of each indicated sgRNA on the growth of Cas9-expressing Ba/F3 cells. Cell growth was computed with respect to cells treated with a non-targeting (control) sgRNA and the percentages of GFP$^{+}$ cells on day 14 were normalized to the percentages on day 2. The illustrated values correspond to the mean computed value over $n = 3$ biological replicates. *Rpa3* sgRNAs were used as a negative control. **e**, As in **d**, but for 32Dcl3 cells. **f**, As in **d**, but for the indicated melanoma and pancreatic ductal adenocarcinoma cells. **g**, As in **d**, but for RN2 cells. **h**, Sequence conservation of the *BRD9* poison exon locus as estimated by phyloP[60]. Conservation and repetitive element annotation is from the UCSC Genome Browser[43]. **i**, RT–PCR analysis of *Brd9* poison exon inclusion using whole bone marrow cells from *Mx1-cre* $Sf3b1^{WT/WT}$ (WT/WT) and *Mx1-cre* $Sf3b1^{K700E/WT}$ (K700E/WT) mice. Three weeks after pIpC treatment, RT–PCR was performed with mouse primers corresponding to those used to assay *BRD9* poison exon inclusion in human cells. Representative images from $n = 2$ technically independent replicates. **j**, RT–PCR analysis to confirm mutant-SF3B1-dependent inclusion of the *BRD9* poison exon in isogenic NALM-6 cell lines engineered to contain the indicated mutations. $SF3B1^{K700K}$ is a wild-type control for genome engineering. Representative images from $n = 2$ technically independent replicates. **k**, Western blot for Flag, SF3B1 and BRD9 in K562 cells overexpressing N-terminally Flag-tagged wild-type *SF3B1* or $SF3B1^{K700E}$ cDNAs, or an empty vector; this panel corresponds to the cells evaluated in **m**. Representative images from $n = 2$ biologically independent replicates. **l**, Western blot for SF3B1 in K562 cells treated with doxycycline-inducible *SF3B1*-targeting shRNAs or a non-targeting control shRNA (shRen); this panel corresponds to cells evaluated in **m**. Representative images from $n = 2$ technically independent replicates. **m**, RT–PCR illustrating the specificity of *BRD9* poison exon inclusion for *SF3B1*-mutated cells in the indicated cell lines. These include K562 cells treated with control shRNA (shRen) or *SF3B1*-targeting shRNAs (the columns labelled 'K562 knock-down'); knock-in of the $SF3B1^{K700K}$, $SF3B1^{K700E}$ or $SF3B1^{K666N}$ mutation into the endogenous locus of *SF3B1* (the columns labelled 'K562 knock-in'); or overexpression of wild-type *SF3B1* or $SF3B1^{K700E}$ cDNA (the columns labelled 'K562 cDNA'). The two right-most lanes show acute myeloid leukaemia cell lines with wild-type *SF3B1* (MV4;11) or a naturally occurring endogenous $SF3B1^{K700E}$ mutation (HNT34 cells; the columns labelled 'leukaemia cell lines'). Representative images from $n = 3$ biologically independent experiments. **n**, As in **m**, but for the indicated pancreatic ductal adenocarcinoma cell lines (left), UVM cell lines (centre) and a cohort of patients with chronic lymphocytic leukaemia (right). CFPAC1 and MIA PaCa2 cells lack *SF3B1* mutations; Panc05;04 cells carry $SF3B1^{Q699H/K700E}$; UPMD1 and MEL270 cells lack *SF3B1* mutations; MEL202 and UPMD2 cells carry $SF3B1^{R625G}$ and $SF3B1^{Y623H}$ mutations, respectively. Sample identifiers for patients with chronic lymphocytic leukaemia correspond to the genotypes shown in Supplementary Table 7. Representative images from $n = 2$ technically independent experiments (left and centre) and $n = 3$ biologically independent experiments (right). **o**, RNA-seq read coverage plots of the *BRD9* poison exon locus from patient samples with the indicated *SF3B1* genotypes. All *SF3B1*-mutated samples exhibit *BRD9* poison exon inclusion. **p**, As in **o**, but for the indicated tissues from healthy donors (from BodyMap 2.0). **q**, qRT–PCR measurement of the half-lives of the poison exon inclusion (left) and exclusion (right) isoforms in isogenic K562 $SF3B1^{K700E}$ cells treated with the indicated shRNAs and actinomycin D to inhibit transcription. NMD inhibition via *UPF1* knockdown stabilizes the inclusion, but not exclusion, isoform. Red arrows indicate primers used to specifically detect the two isoforms. $n = 2$ biologically independent experiments and $n = 2$ technically independent experiments for the inclusion isoform; $n = 3$ technically independent experiments for the exclusion isoform. $P$ value was calculated by two-sided $t$-test at 8 h. **r**, Bar graph illustrating the estimated poison exon inclusion isoform half-life in the indicated conditions from the data in **q**. Error bars, mean + s.d. $n = 2$ biologically independent experiments and $n = 2$ technically independent experiments. $P$ value was calculated by two-sided $t$-test. **s**, As in **r**, but for the exclusion isoform. Error bars, mean + s.d. $n = 3$ technically independent experiments. $P$ value was calculated by two-sided $t$-test. **t**, As in **q**, but for NALM-6 $SF3B1^{K700E}$ cells. $n = 3$ technically independent experiments for the inclusion isoform and the exclusion isoform. $P$ value was calculated by two-sided $t$-test at 8 h. **u**, As in **r**, but for NALM-6 $SF3B1^{K700E}$ cells. $n = 3$ technically independent experiments. Error bars, mean + s.d. $P$ value was calculated by two-sided $t$-test. **v**, As in **s**, but for NALM-6 $SF3B1^{K700E}$ cells. Error bars, mean + s.d. $n = 3$ technically independent experiments. $P$ value was calculated by two-sided $t$-test. **w**, Western blot for BRD9 in NALM-6 cells with or without knock-in of an *SF3B1* mutation. Actin, loading control. Representative images from $n = 3$ biologically independent experiments. **x**, Rank plot of *BRD9* poison exon inclusion (scale of 0 to 1; top) and box plot of gene expression (inset) for patients stratified by *SF3B1* mutational status (data are from cohorts of patients with myelodysplastic syndromes or UVM, as in Fig. 1a). *SF3B1* mutations were strongly associated with high poison exon inclusion and low *BRD9* expression. Boxes illustrate 1st and 3rd quartiles, with whiskers extending to 1.5× interquartile range. $P$ value computed with one-sided Mann–Whitney $U$-test.

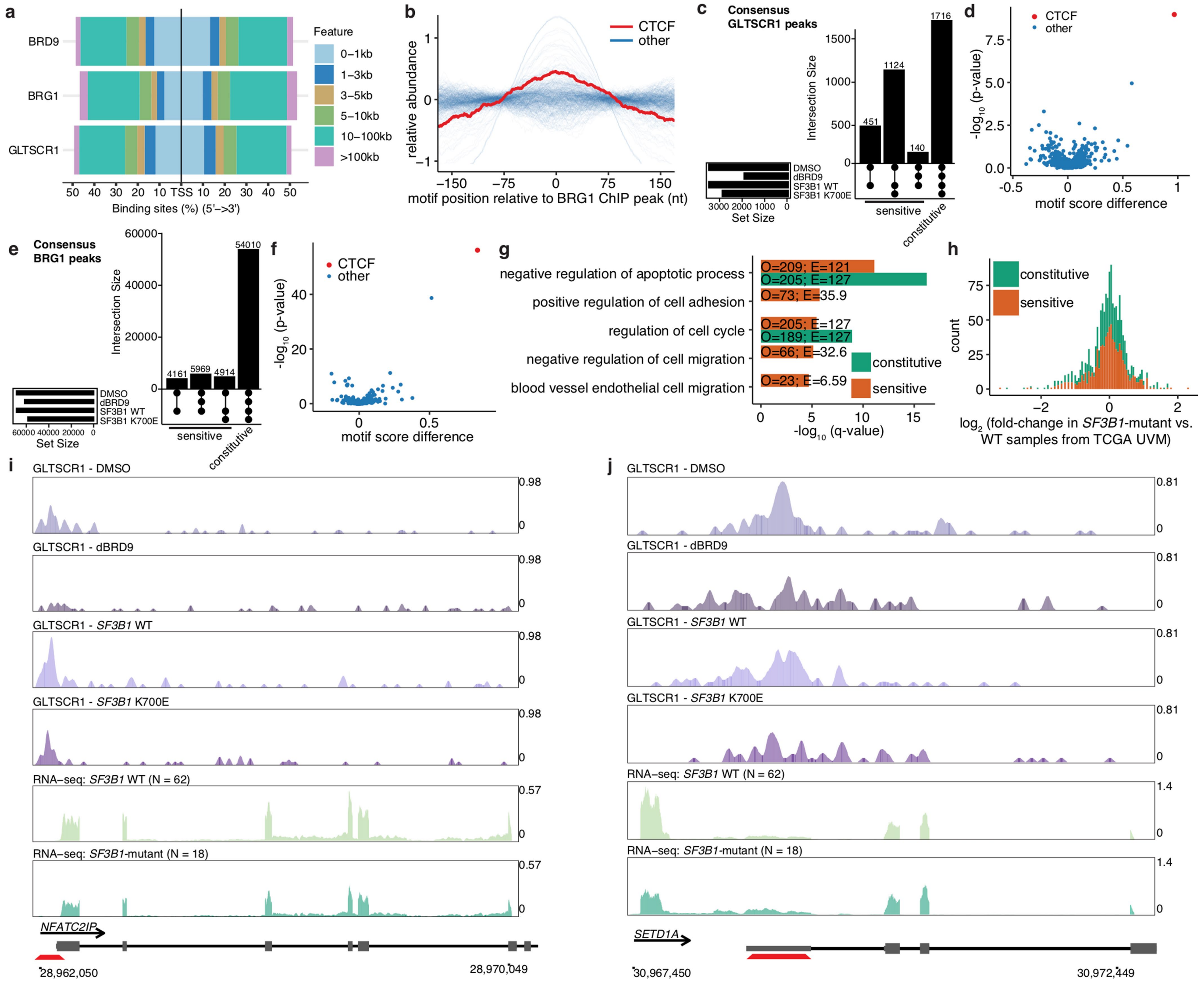


**Extended Data Fig. 2 |** See next page for caption.

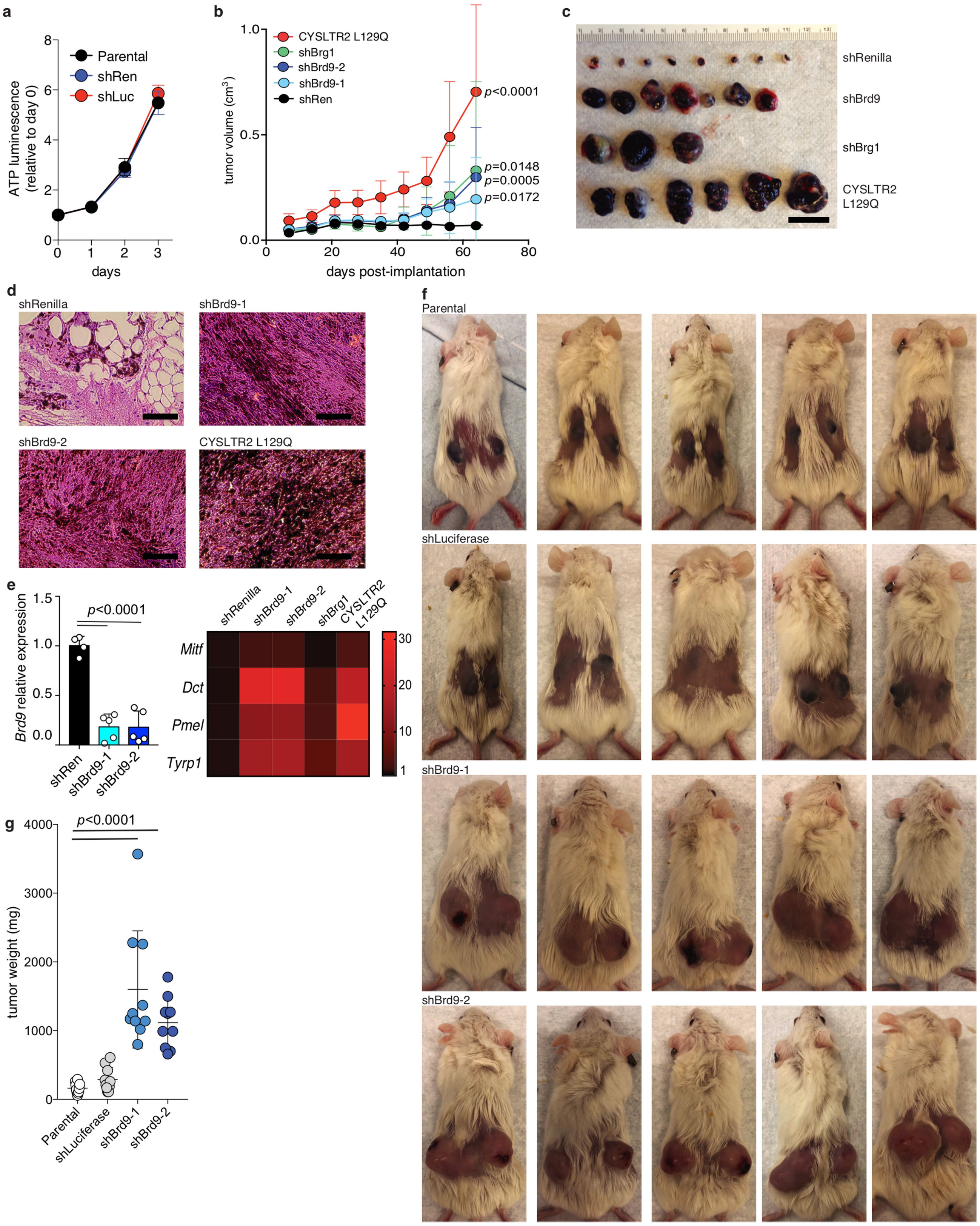
**Extended Data Fig. 2 | Mutant SF3B1 recognizes an aberrant branchpoint within *BRD9* to promote poison exon inclusion, causing loss of full-length BRD9 protein. a**, Schematic illustrating the strategy for knock-in of an HA tag into the endogenous *BRD9* locus. The single-stranded donor DNA contained a 197-nt fragment, including 83 nt homologous to the *BRD9* 5′ UTR (upstream of the HA tag) and 87 nt homologous to *BRD9* exon 1 (downstream of the start codon). **b**, Sanger sequencing of genomic DNA validating successful HA tag knock-in in K562 *SF3B1*$^{K700E}$ cells. Representative images from $n = 2$ biologically independent experiments. **c**, Western blot with anti-BRD9 (left), anti-HA (right, top) or anti-actin (right, bottom) used to probe K562 *SF3B1*$^{K700E}$ cells carrying an endogenously N-terminally HA-tagged BRD9. White arrows, non-specific bands. Red arrow, expected size of BRD9 protein. Representative images from $n = 3$ biologically independent experiments. **d**, Western blot for Flag, SF3B1 and endogenous BRD9 protein in MEL270 cells with doxycycline-inducible Flag-tagged wild-type SF3B1 or Flag-tagged SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments. **e**, Sanger sequencing of cDNA arising from reverse transcription of lariats arising from inclusion (top) (exon 14–exon 14a splicing) or exclusion (bottom) (exon 14–exon 15 splicing) of the *BRD9* poison exon in MEL270 cells with doxycycline-inducible Flag-tagged wild-type SF3B1 (bottom) or Flag-tagged SF3B1(K700E) (top). The branchpoints are illustrated in Fig. 2a. Representative images from $n = 3$ biologically independent experiments. **f**, As in **e**, but for T47D cells. Representative images from $n = 3$ biologically independent experiments. **g**, As in Fig. 2b, but for the indicated minigene mutagenesis in T47D cells with doxycycline-inducible Flag-tagged SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments. **h**, Western blot of U2AF2, U2AF1 and histone H3 in K562 cells transfected with siRNAs against *U2AF1* and/or *U2AF2* (top) and bar plot illustrating mean *BRD9* poison exon inclusion as measured by quantitative PCR (qPCR) following siRNA knockdown of *U2AF1* and/or *U2AF2* (bottom). Experiment performed with $n = 1$ biologically independent replicate for siRNA transfection, $n = 1$ technically independent replicate for western blot and $n = 3$ technically independent replicates for RT–PCR. Poison exon inclusion was computed over all $n = 3 \times 3$ (9) combinations of technical replicates for RT–PCR for the inclusion and exclusion isoforms. Bars illustrate mean inclusion. **i**, *EPB49* cassette exon inclusion as measured by qPCR following siRNA knockdown of *U2AF1* and/or *U2AF2*. As the *EPB49* cassette exon is U2AF-dependent, this experiment serves as a positive control for the functional efficacy of *U2AF1* and *U2AF2* knockdown. $n = 3$ technically independent experiments. Cassette exon inclusion was computed over all $n = 3 \times 3$ (9) combinations of technical replicates for RT–PCR for the inclusion and exclusion isoforms. Bars illustrate mean inclusion. **j**, As in Fig. 2b, but for the indicated minigene mutagenesis in T47D cells with doxycycline-inducible Flag-tagged SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments. **k**, As in Fig. 2c, but for the indicated minigene mutagenesis in T47D cells with doxycycline-inducible Flag-tagged SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments. **l**, Western blot for Flag, SF3B1, BRD9 and actin in MEL270 cells expressing an empty vector or N-terminally Flag-tagged wild-type *SF3B1*, *SF3B1*$^{R625H}$ or *SF3B1*$^{K700E}$ cDNA. Representative images from $n = 3$ biologically independent experiments. **m**, RT–PCR analysis of *BRD9* splicing in MEL270 cells expressing doxycycline-inducible empty vector, wild-type SF3B1, SF3B1(R625H) or SF3B1(K700E). The left column illustrates minigene splicing and the right column illustrates endogenous *BRD9* splicing. Representative images from $n = 3$ biologically independent experiments. **n**, As in **m**, but for the illustrated minigene mutations at the 5′ end of the poison exon. Representative images from $n = 3$ biologically independent experiments. **o**, Mutations generated at the 5′ end of the *BRD9* poison exon by CRISPR–Cas9-mediated indels in MEL202 cells (*SF3B1*$^{R625G}$). The PAM sequence is illustrated with uppercase, underlined nucleotides. Red nucleotides hybridize to the sgRNA. Substitutions are illustrated with lowercase, underlined nucleotides. **p**, As in **o**, but for MEL270 cells. Representative images from $n = 3$ biologically independent experiments. **q**, As in Fig. 2d top, but for MEL270 cells. Representative images from $n = 3$ biologically independent experiments. **r**, As in Fig. 2d bottom, but for MEL270 cells. Representative images from $n = 3$ biologically independent experiments.



**Extended Data Fig. 3 | BRD9 loss impairs ncBAF complex formation.** **a**, Mutation rate observed across TCGA cohorts for canonical BAF, polybromo-associated BAF and ncBAF components. **b**, Western blot confirming Flag-tagged BRD9 protein expression in 3 × Flag–BRD9-expressing K562 cells. Representative images from $n = 3$ biologically independent experiments. **c**, Experimental workflow for using rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME)[33] for purification and identification of the chromatin-associated interactions partners of BRD9. **d**, Cross-linking and immunoprecipitation with IgG or Flag followed by probing with the indicated antibodies. Data from 3 × Flag–BRD9-expressing MEL270 cells. Representative images from $n = 3$ biologically independent experiments. **e**, Immunoprecipitation of GLTSCR1 followed by western blotting with the indicated antibodies in *SF3B1*$^{K700E}$ knock-in NALM-6 cells. Representative images from $n = 3$ biologically independent experiments. **f**, Immunoprecipitation of GLTSCR1 (top) or BRG1 (bottom) followed by blotting with the indicated antibodies in K562 cells with CRISPR-mediated knockout (KO) of BRD9. Representative images from $n = 3$ biologically independent experiments. **g**, Schematic of the BRD9 full-length (FL) protein and the deletion mutants that we constructed. BD, bromodomain; DUF, domain of unknown function; N, amino acids 1–133 of BRD9; N + BD, amino acids 1–242 of BRD9; N + BD + DUF, amino acids 1–505 of BRD9; dN, amino acids 134–597 of BRD9; dBD, bromodomain-deletion mutant of BRD9; dDUF, DUF-domain deletion mutant of BRD9. **h**, Immunoprecipitation with Flag following by probing (immunoblot, IB) for GLTSCR1 or GLTSCR1L in 293T cells expressing 3 × Flag-tagged versions of the indicated deletion mutants. Deletion mutants illustrated in **g**. Representative images from $n = 3$ biologically independent experiments.

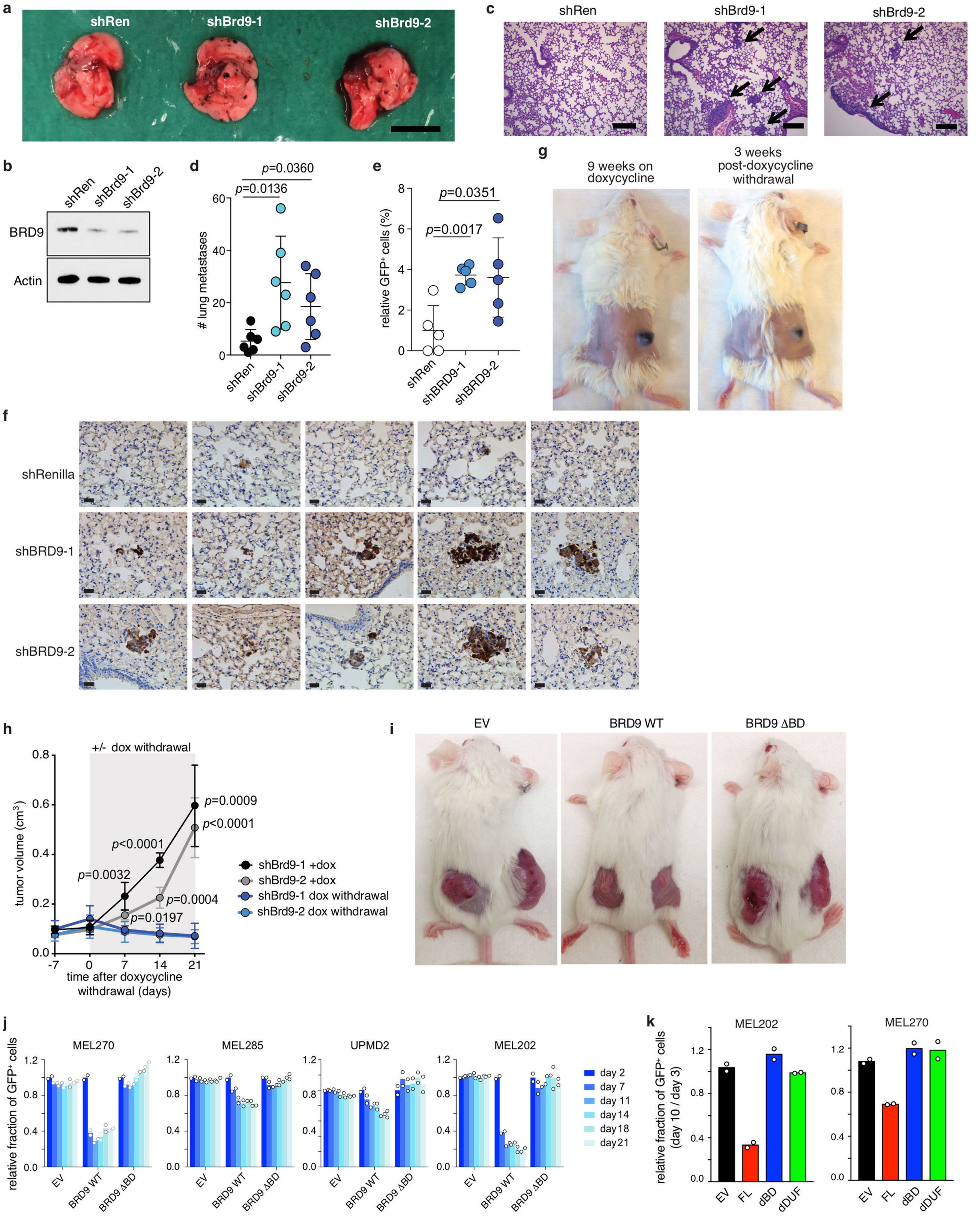


**Extended Data Fig. 4 | BRD9 loss drives relocalization of GLTSCR1 away from CTCF-associated loci. a**, As in Fig. 3e, but illustrating relative positions with respect to transcription start sites (TSSs). **b**, As in Fig. 3f, but for motifs at BRG1-bound loci. *n* = 401 transcription factors analysed. **c**, UpSet plots depicting the overlap of consensus GLTSCR1-bound loci in MEL270 cells with the indicated treatments. **d**, Volcano plot illustrating the difference in the mean motif scores at BRD9-sensitive versus constitutive GLTSCR1-bound loci for the transcription factors in Fig. 3f, as well as associated statistical significance. *n* = 401 transcription factors analysed. *P* values computed with a two-sided Mann–Whitney *U*-test. **e**, As in **c**, but for BRG1-bound loci. **f**, As in **d**, but for BRG1-bound loci. *n* = 401 transcription factors analysed. **g**, Selected enriched annotation terms from a Genomic Regions Enrichment of Annotations Tool (GREAT) analysis[61] of genes near BRD9-sensitive and constitutive GLTSCR1-bound loci. Plot illustrates $-\log_{10}$(FDR), computed with a one-sided binomial test and corrected for multiple testing using the Benjamini–Hochberg procedure. O and E, numbers of genes that were observed and expected, respectively. **h**, Differences in gene expression in *SF3B1*-mutated versus wild-type samples in the TCGA UVM cohort for genes with GLTSCR1-bound promoters identified in MEL270 cells. Colours indicate the responsiveness of peaks to BRD9 loss. **i**, Read coverage from GLTSCR1 ChIP–seq (MEL270 cells) and RNA-seq (TCGA UVM cohort) around *NFATC2IP*. Red trapezoid indicates GLTSCR1 binding in the promoter, with reduced binding upon treatment with BRD9 degrader or expression of $SF3B1^{K700E}$. *NFATC2IP* was significantly differentially expressed in UVM samples with *SF3B1* mutations relative to wild-type samples. Vertical axis scales were rendered comparable by normalizing ChIP–seq read coverage to mapped library size and RNA-seq read coverage to mapped library size, restricted to coding genes. ChIP–seq experiment performed for *n* = 1 biologically independent replicate. **j**, As in **i**, but for *SETD1A*. *SETD1A* was significantly differentially expressed in UVM samples with *SF3B1* mutations relative to wild-type samples.



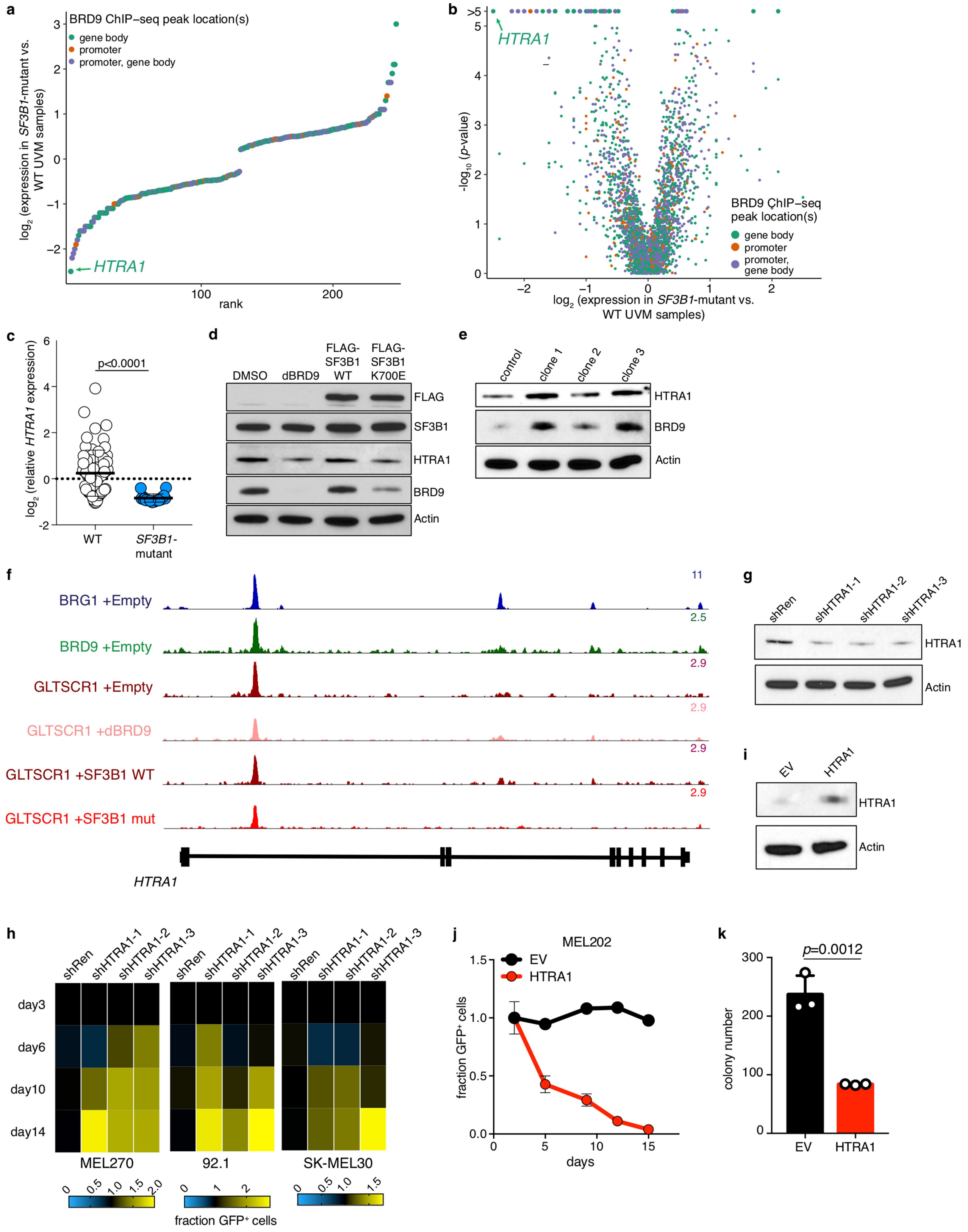
**Extended Data Fig. 5** | See next page for caption.

**Extended Data Fig. 5 | BRD9 is a potent tumour suppressor in UVM. a**, In vitro growth curves of Melan-a cells treated with two non-targeting shRNAs (against *Renilla* (shRen) and luciferase (shLuc)) versus parental, un-manipulated Melan-a cells. $n = 3$ biologically independent experiments per group. Data are presented as mean ± s.d. **b**, Tumour volume in SCID mice subcutaneously injected with Melan-a cells expressing a control shRNA (against *Renilla*), shRNA against *Brd9* (*Brd9* shRNA no. 1 and no. 2), shRNA against *Brg1* or cDNA encoding CYSLTR2(L129Q) ($n = 8$ mice per group). Data are presented as mean ± s.d. $P$ value at day 64 was calculated compared to *Renilla* shRNA group with a two-sided $t$-test. **c**, Representative images of the dissected melanomas from **b**. **d**, H&E images of melanomas from **b**. Scale bars, 100 μm. Representative images from $n = 3$ biologically independent experiments. **e**, qRT–PCR measuring expression of *Brd9* (left) and melanoma-associated genes (*Mitf*, *Dct*, *Pmel* and *Tyrp1*) of melanomas from **a**. $n = 4$ (*Renilla* shRNA) and $n = 5$ (*Brd9* shRNA no. 1 and no. 2) biologically independent experiments. Data are presented as mean ± s.d. $P$ value was calculated by two-sided $t$-test. **f**, Images of mice transplanted with parental, un-manipulated Melan-a cells or Melan-a cells transduced with a non-targeting shRNA or *Brd9*-targeting shRNA. Cells were subcutaneously engrafted into SCID mice and tumour volume was estimated 36 days after transplant. **g**, Volumes of tumours from **f** at day 36. Data are presented as mean ± s.d. $n = 10$ tumours per group. $P$ value was calculated relative to the parental group by a two-sided $t$-test.



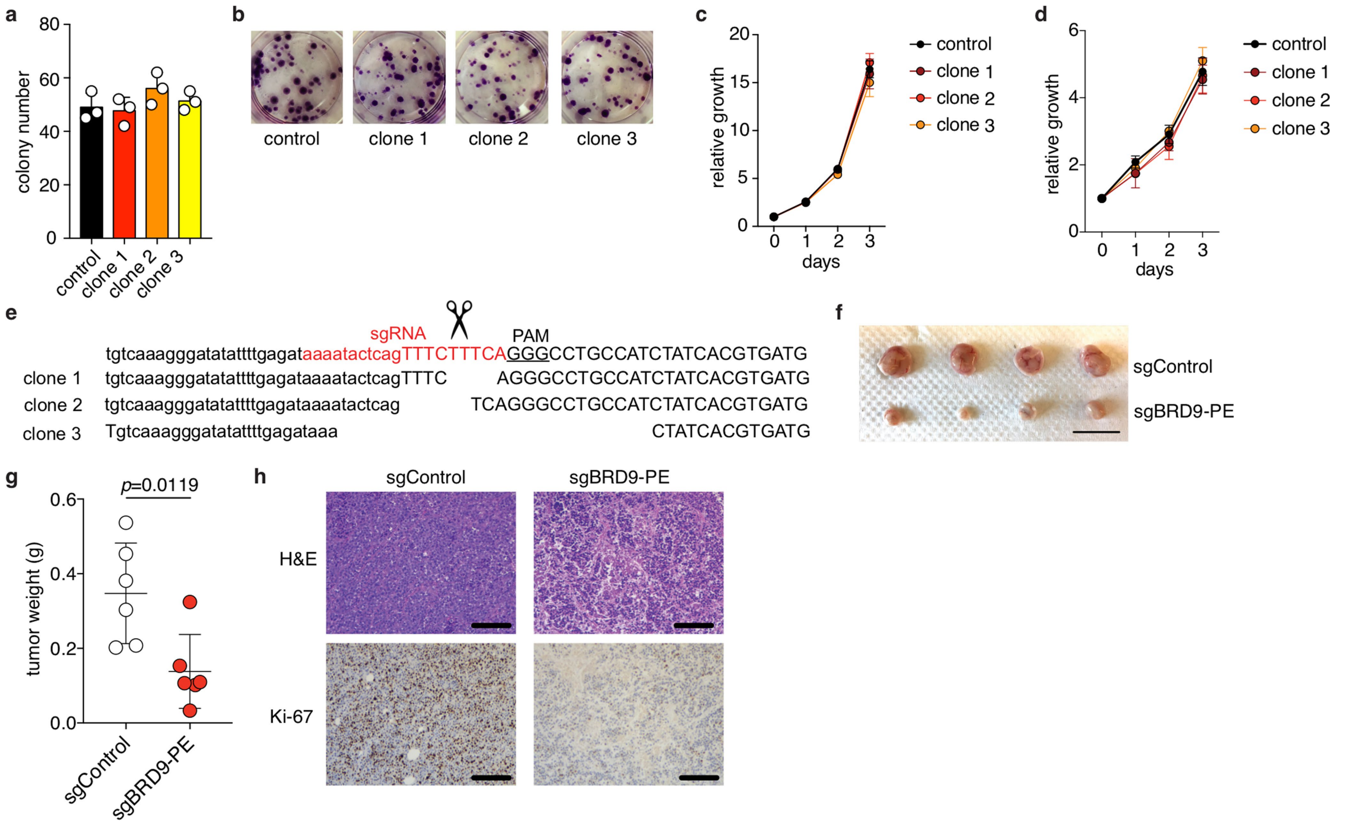
**Extended Data Fig. 6** | See next page for caption.

**Extended Data Fig. 6 | BRD9 is a potent tumour suppressor in UVM.** **a**, Representative images of pulmonary metastatic foci produced 14 days after intravenous injection of B16 cells with or without *Brd9* shRNA (MLS-E vector). Scale bar, 5 mm. **b**, Western blot of endogenous BRD9 in B16 cells immediately before injection. Actin, loading control. The experiment was repeated three times with similar results. **c**, H&E sections of lung metastases. Arrows indicate metastatic foci. Scale bar, 100 μm. The experiment was repeated three times with similar results. **d**, Numbers of pulmonary B16 metastases identified in the experiments from **a**. $n = 6$ mice per group. $P$ value was calculated relative to the *Renilla* shRNA group by a two-sided $t$-test. **e**, Relative percentages of $GFP^+$ 92.1 cells with or without *BRD9* shRNA (MLS-E vector), assessed by flow cytometric analysis of lung tissue in recipient NSG (NOD–SCID $Il2rg^{-/-}$) mice 14 days after intravenous injection by tail vein. The signal was normalized by dividing by the average percentage of $GFP^+$ cells in the *Renilla* shRNA (control) group. $n = 5$ biologically independent experiments per group. $P$ value was calculated relative to the *Renilla* shRNA group by a two-sided $t$-test. **f**, Anti-GFP immunohistochemistry for sections of lung metastases from the experiment in **e**. Scale bar, 200 μm. The experiment was repeated three times with similar results. **g**, Representative images of tumours derived from transplantation of Melan-a cells transduced with doxycycline-inducible *Brd9* shRNA. Doxycycline was administered for nine weeks (left) and followed by doxycycline withdrawal for three weeks (right). **h**, Tumour volume for the experiment in **g**. $n = 4$ mice per group. The experiment was repeated twice with similar results. $P$ value was calculated relative to the parental group by a two-sided $t$-test at day 7, day 14 and day 21. **i**, Representative images of recipient mice engrafted with MEL270 cells transduced with empty vector, full-length BRD9 (BRD9 WT) or a bromodomain-deletion mutant of BRD9 (BRD9 ΔBD) at day 12. $n = 5$ mice per group. **j**, Results of a competition assay to measure the effects of expression of the indicated cDNAs on growth of the indicated melanoma cells. Transduced cells were identified by co-expression of GFP (pMIGII vector). The percentage of $GFP^+$ cells was tracked over 21 days and normalized to the GFP percentage on day 2. Data are presented as mean ± s.d. $n = 2$ biologically independent experiments per group. **k**, Results of a competition assay to measure the effects of expression of the indicated cDNAs on growth of the indicated melanoma cells. Transduced cells were identified by the co-expression of GFP (pMIGII vector). The percentage of $GFP^+$ cells was tracked over 10 days and normalized to the GFP percentage on day 3. $n = 2$ biologically independent experiments per group.

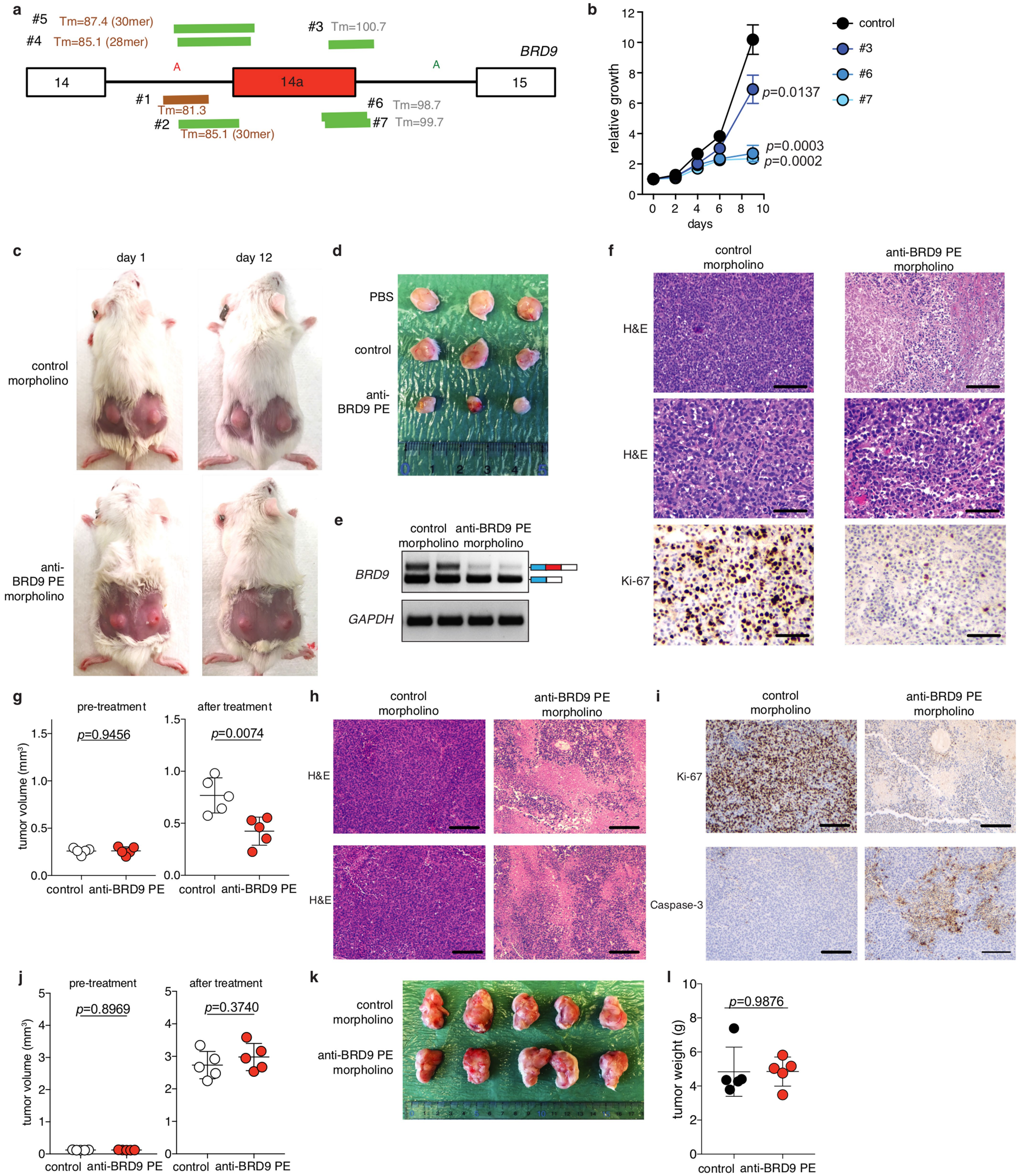


**Extended Data Fig. 7** | See next page for caption.

**Extended Data Fig. 7 | BRD9 regulates *HTRA1* expression to promote UVM tumorigenesis. a**, Rank plot illustrating fold change (expressed in $\log_2$) of each significantly differentially expressed gene identified by comparing samples from patients in the TCGA UVM cohort with mutated versus wild-type *SF3B1*. The plot is restricted to genes with BRD9 ChIP–seq peaks within their promoters or gene bodies in the absence of perturbations to BRD9 (MEL270 cells treated with DMSO or following ectopic expression of wild-type SF3B1). $n = 3{,}122$ genes analysed, of which $n = 248$ met the significance ($P < 0.001$) and expression (median expression in both wild-type and mutant samples > 2 transcripts per million) thresholds, and so are illustrated here. *P* value was computed using a two-sided Mann–Whitney *U*-test. **b**, As in **a**, but a volcano plot additionally illustrating the *P* value associated with the comparison between *SF3B1*-mutated and wild-type *SF3B1* samples. $n = 3{,}122$ genes analysed and illustrated. *P* value computed with two-sided Mann–Whitney *U*-test. **c**, *HTRA1* expression in samples from patients in the TCGA UVM cohort with ($n = 18$) or without ($n = 62$) *SF3B1* mutations. Expression is *z*-score normalized across all samples. Data are presented as mean ± s.d. *P* value computed with two-sided *t*-test. **d**, Western blot for Flag, SF3B1, HTRA1, BRD9 and actin in MEL270 cells (wild-type *SF3B1*), treated with DMSO, BRD9 degrader, Flag–SF3B1(WT) or Flag–SF3B1(K700E). Representative images from $n = 3$ biologically independent experiments. **e**, Western blot for HTRA1, BRD9 and actin in MEL202 cells ($SF3B1^{R625G}$) following CRISPR–Cas9-mediated mutagenesis of the *BRD9* poison exon (as shown in Extended Data Fig. 2o). Representative images from $n = 3$ technically independent experiments. **f**, Read coverage for BRG1, BRD9 and GLTSCR1 ChIP–seq at the *HTRA1* locus in MEL270 cells (shown in **d**) treated with an empty vector, BRD9 degrader, or $SF3B1^{WT}$ or $SF3B1^{K700E}$ cDNAs ($n = 1$ ChIP–seq experiment performed for each condition). **g**, Western blot for HTRA1 and actin in MEL270 cells treated with shRNAs against *HTRA1* or with a non-targeting control shRNA (against *Renilla*). Representative images from $n = 3$ biologically independent experiments. **h**, Heat map summarizing the results of a competition assay to measure the effect of each indicated shRNA on the growth of Cas9-expressing UVM cell lines with wild-type *SF3B1*. Cell growth was computed with respect to cells treated with a non-targeting control shRNA (against *Renilla*) and the percentage of $GFP^+$ cells at day 14 was normalized to that at day 2. The illustrated values correspond to the mean values computed over $n = 3$ biologically independent experiments. **i**, Western blot for HTRA1 and actin in MEL202 cells ($SF3B1^{R625G}$) following stable overexpression of an empty vector or HTRA1 (both with an MSCV-IRES-GFP vector). Representative images from $n = 3$ biologically independent experiments. **j**, Ratio of $GFP^+$ to $GFP^-$ MEL202 cells ($SF3B1^{R625G}$) from a competition experiment in which $GFP^+$ cells from **i** were seeded at an initial 1:1 ratio with $GFP^-$ control cells. Data are presented as mean of $n = 2$ biologically independent experiments. **k**, Colony number of MEL202 cells expressing empty vector or *HTRA1* cDNA from **i**, following 10 days of growth in soft agar. Data are presented as mean of $n = 3$ biologically independent experiments.

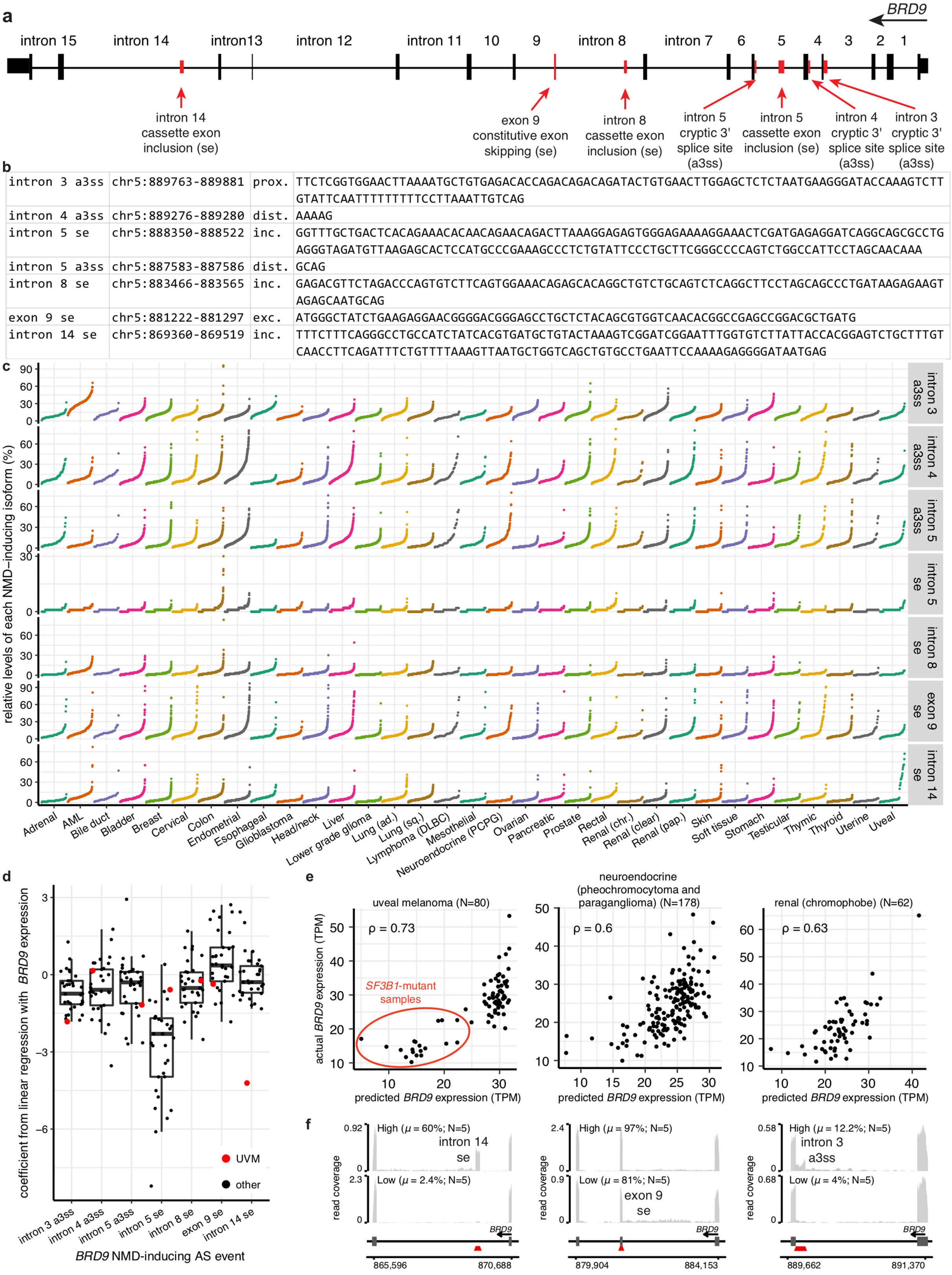


**Extended Data Fig. 8 | CRISPR–Cas9-mediated mutagenesis of the *BRD9* poison exon corrects *BRD9* aberrant splicing and abrogates growth of *SF3B1*-mutated melanoma. a**, Colony number for MEL270 cells (wild-type *SF3B1*) without (control) or with (clone 1, clone 2 and clone 3) CRISPR–Cas9-induced indels that disrupted *BRD9* poison exon recognition. Data presented as mean ± s.d. $n = 3$ biologically independent experiments per group. **b**, Representative images from **a**. **c**, Proliferation of the clones described in **a**. $n = 3$ biologically independent experiments per group. **d**, Proliferation of MEL285 cells (wild-type *SF3B1*) without (control) or with (clones 1, 2 and 3) CRISPR–Cas9-induced indels that disrupted *BRD9* poison exon recognition. $n = 3$ biologically independent experiments per group. **e**, Mutations generated at the 5′ end of the *BRD9* poison exon by CRISPR–Cas9-mediated indels in clones 1, 2 and 3 of MEL285 cells from **d**. The PAM sequence is illustrated with uppercase, underlined nucleotides. Red nucleotides hybridize to the sgRNA. **f**, Representative images of dissected tumours from recipient mice transplanted with CRISPR–Cas9-modified MEL202 clones. **g**, Tumour weight for the tumours illustrated in **f**. Data presented as mean ± s.d. $n = 6$ biologically independent experiments per group. $P$ value was calculated relative to the control shRNA group by a two-sided $t$-test. **h**, H&E staining, as well as Ki-67 immunohistochemistry images, for the tumours illustrated in **f**. Representative images from $n = 3$ independent histological analyses.



**Extended Data Fig. 9** | See next page for caption.

**Extended Data Fig. 9 | Correcting *BRD9* mis-splicing in *SF3B1*-mutated xenografts with ASOs suppresses tumour growth. a**, Cartoon representation of the *BRD9* loci targeted by each designed morpholino. Melting temperature ($T_m$) is shown. Lengths of target sequences are indicated in parentheses; if these are not indicated, then the length is 25 nt. **b**, Growth of MEL202 cells ($SF3B1^{R625G}$) treated with 10 μM of control non-targeting (control) or *BRD9* poison-exon-targeting morpholinos (no. 3, no. 6 and no. 7). $n = 3$ biologically independent experiments per group. *P* values at day 9 were calculated relative to the control group by a two-sided *t*-test. **c**, Representative images of recipient mice xenografted with MEL202 cells and treated with PBS or morpholinos in vivo. Each tumour was analysed after in vivo treatment with PBS, control morpholino or morpholino against the *BRD9* poison exon (no. 6) (12.5 mg $kg^{-1}$, every other day to a total of 8 intratumoral injections). $n = 10$ tumours per group. **d**, Representative images of dissected tumours from the experiment described in **c**. **e**, RT–PCR results of tumours from **c** to evaluate *BRD9* splicing. The experiment was repeated three times with similar results. **f**, Representative H&E and Ki-67 staining images of tumours from **c**. Scale bars, 100 μm (top), 50 μm (middle and bottom). The experiment was repeated three times with similar results. **g**, Estimated tumour volume for recipient mice transplanted with a patient-derived xenograft model of $SF3B1^{R625C}$ rectal melanoma and treated with in vivo morpholinos (control or morpholino against *BRD9* poison exon, no. 6) (12.5 mg $kg^{-1}$, every other day to a total of 8 intratumoral injections). $n = 5$ mice per group. Estimated tumour volumes before and after treatment are shown. Data are presented as mean ± s.d. *P* values were calculated relative to the control group by a two-sided *t*-test. **h**, Representative H&E staining images of tumours from **g**. The experiment was repeated three times with similar results. **i**, Representative Ki-67 staining images of tumours from **g**. The experiment was repeated three times with similar results. **j**, Estimated tumour volume for recipient mice transplanted with a patient-derived xenograft model of UVM (wild-type *SF3B1*), treated with in vivo morpholinos (control or morpholino against *BRD9* poison exon, no. 6) (12.5 mg $kg^{-1}$, every other day to a total of 8 intratumoural injections). $n = 5$ mice per group. Estimated tumour volumes before and after treatment are shown. Data are presented as mean ± s.d. *P* values weres calculated relative to the control group by a two-sided *t*-test. **k**, Representative images of dissected tumours from **j**. **l**, Tumour weight for tumours from **k**. $n = 5$ mice per group. *P* value was calculated relative to the control group by a two-sided *t*-test.



|  |  |  |  |
|---|---|---|---|
| intron 3 a3ss | chr5:889763-889881 | prox. | TTCTCGGTGGAACTTAAAATGCTGTGAGACACCAGACAGACAGATACTGTGAACTTGGAGCTCTCTAATGAAGGGATACCAAAGTCTTGTATTCAATTTTTTTTTCCTTAAATTGTCAG |
| intron 4 a3ss | chr5:889276-889280 | dist. | AAAAG |
| intron 5 se | chr5:888350-888522 | inc. | GGTTTGCTGACTCACAGAAACACAACAGAACAGACTTAAAGGAGAGTGGGAGAAAAGGAAACTCGATGAGAGGATCAGGCAGCGCCTGAGGGTAGATGTTAAGAGCACTCCATGCCCGAAAGCCCTCTGTATTCCCTGCTTCGGGCCCCAGTCTGGCCATTCCTAGCAACAAA |
| intron 5 a3ss | chr5:887583-887586 | dist. | GCAG |
| intron 8 se | chr5:883466-883565 | inc. | GAGACGTTCTAGACCCAGTGTCTTCAGTGGAAACAGAGCACAGGCTGTCTGCAGTCTCAGGCTTCCTAGCAGCCCTGATAAGAGAAGTAGAGCAATGCAG |
| exon 9 se | chr5:881222-881297 | exc. | ATGGGCTATCTGAAGAGGAACGGGGACGGGAGCCTGCTCTACAGCGTGGTCAACACGGCCGAGCCGGACGCTGATG |
| intron 14 se | chr5:869360-869519 | inc. | TTTCTTTCAGGGCCTGCCATCTATCACGTGATGCTGTACTAAAGTCGGATCGGAATTTGGTGTCTTATTACCACGGAGTCTGCTTTGTCAACCTTCAGATTTCTGTTTTAAAGTTAATGCTGGTCAGCTGTGCCTGAATTCCAAAAGAGGGGATAATGAG |

**Extended Data Fig. 10 |** See next page for caption.

**Extended Data Fig. 10 | Use of multiple, distinct NMD isoforms of *BRD9*.** **a**, *BRD9* gene structure illustrating constitutive *BRD9* exons and alternative splicing events that are predicted to induce NMD. *SF3B1* mutations promote inclusion of the BRD9 poison exon in intron 14. **b**, Genomic coordinates (hg19/GRCh37 assembly) of each NMD-inducing event illustrated in **a**, as well as genomic sequence of each alternatively spliced region highlighted in red in **a**. The third column indicates the specific isoform that is a predicted NMD substrate. Prox, intron-proximal competing 3′ splice site; dist, intron-distal competing 3′ splice site; inc, exon inclusion; exc, exon exclusion. **c**, Rank plot illustrating levels of each NMD-inducing isoform relative to total *BRD9* mRNA levels for each sample in each indicated TCGA cohort. Boxes illustrate first and third quartiles, with whiskers extending to 1.5× interquartile range. **d**, Box plot illustrating the distribution of coefficients estimated by fitting a linear model to predict *BRD9* gene expression on the basis of relative levels of each NMD-inducing isoform. The relative levels of NMD-inducing isoforms illustrated in **c**, as well as *BRD9* gene expression estimates for each sample, were used to construct an independent linear model with robust regression for each TCGA cohort. The coefficients resulting from this model fitting procedure are illustrated in the box plot, in which each dot corresponds to the coefficient associated with the corresponding NMD-inducing event for a single TCGA cohort. Coefficients for the TCGA UVM cohort are highlighted in red. The coefficients are typically negative (as expected for NMD-inducing isoforms), with the exception of constitutive exon 9 skipping, for which the coefficients are generally positive—as expected for an event in which NMD is induced when a constitutive exon is excluded. The *SF3B1*-mutation-responsive poison exon in intron 14 dominates the fit for UVM, as expected. $n = 33$ TCGA cohorts analysed and illustrated. **e**, Scatter plots comparing actual ($y$ axis) and predicted ($x$ axis) *BRD9* expression levels for three TCGA cohorts. Each dot corresponds to a single sample. $\rho$, Spearman's correlation between actual and predicted values. **f**, RNA-seq read coverage plots for patient samples from the TCGA cohorts illustrated in **e** for representative alternative splicing events illustrated in **a**. Each coverage plot illustrates data averaged over the $n = 5$ patient samples from the vertically matched cohort in **e** that exhibit the lowest or highest relative expression of the NMD-inducing isoform. $\mu$, mean relative expression of the illustrated NMD-inducing isoform, computed over each group of samples.

nature research

Corresponding author(s): Omar Abdel-Wahab
Robert K. Bradley

Last updated by author(s): Jul 6, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☐ | ☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was used for data collection. |
| Data analysis | Publicly available software was used in this study. Specific programs are RSEM (v1.2.4), Bowtie (v1.0.0), TopHat (v2.0.8b), MISO (v2.0), and Bioconductor (v3.7) within the R (v3.5.1) programming environment. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq and ChIP-seq data generated as part of this study were deposited in the Gene Expression Omnibus (accession number GSE124720). RNA-seq data from published studies were downloaded from CGHub (TCGA UVM57), EMBL-EBI ArrayExpress (Illumina Human BodyMap 2.0: E-MTAB-513), the Gene Expression Omnibus (accession numbers GSE72790, GSE114922) (CLL: Darman et al. 201515, MDS: Pellagati et al. 201858), or directly obtained from the authors (UVM: Alsafadi et al. 201610). Gel source data can be found in Supplementary Fig. 1. Other data that support this study’s findings are available from the authors upon reasonable request.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample sizes for xenograft experiments were chosen based on published studies of known oncogenic drivers of relevant models (e.g., expression of the oncoprotein CYSLTR2 L129Q in Melan-a cells). |
| Data exclusions | No data were excluded. |
| Replication | Attempts at replication were successful. |
| Randomization | Animals were randomly assigned to experimental groups. |
| Blinding | The data presented did not require the use of blinding. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | Antibodies |
| ☐ | ☒ | Eukaryotic cell lines |
| ☒ | ☐ | Palaeontology |
| ☐ | ☒ | Animals and other organisms |
| ☐ | ☒ | Human research participants |
| ☒ | ☐ | Clinical data |

## Methods

| n/a | Involved in the study | |
|---|---|---|
| ☐ | ☒ | ChIP-seq |
| ☒ | ☐ | Flow cytometry |
| ☒ | ☐ | MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | For Western blotting, the following antibodies to the following proteins were used: BRD9 (Bethyl Laboratories; A303-781A and Active Motif; 61538), SF3B1/Sap-155 (MBL; D221-3), Flag-M2 (Sigma-Aldrich; F-1084), β-actin (Sigma-Aldrich; A-5441), GLTSCR1 (Santa Cruz Biotechnology; sc-515086), GLTSCR1L (Thermo Fisher Scientific; PA5-56126), BRM (Bethyl Laboratories; A303-015A), BRG1 (Santa Cruz Biotechnology; sc-17796), BAF155 (Santa Cruz Biotechnology; sc-48350), BAF60A (Santa Cruz Biotechnology; sc-135843), BAF47 (Santa Cruz Biotechnology; sc-166165), ARID1A (Santa Cruz Biotechnology; sc-373784), ARID2 (Santa Cruz Biotechnology; sc-166117), BRD7 (Thermo Fisher Scientific; PA5-49379), U2AF2 (Bethyl Laboratories; A303-665A), U2AF1 (Bethyl Laboratories; A302-080A), Histone H3 (Abcam; ab1791), HTRA1 (R&D systems; MAB2916-SP). For ChIP-seq studies in MEL270 cells, antibodies to endogenous BRG1 (Abcam EPNCIR111A, Lot # GR3208604-8), GLTSCR1 (Santa Cruz SC-240516, Lot # A2313), and BRD9 (Abcam, ab137245) were used. |
| Validation | All antibodies were validated by the supplier for human samples, and were checked in the lab by Western blotting on cell lysate and by comparing to the manufacturer's or in-house results. |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | Ba/F3, K562, NALM6, TF1, HEK293T, MCF10A, Miapaca2, CFPAC1, B16, and HEK293T cells were obtained from the American Type Culture Collection (ATCC) . The isogenic K562 and NALM6 cell lines with or without SF3B1 mutations were generated at Horizon Discovery. Melan-a cells were provided by Dr. Dorothy Bennett (PMID: 3102392); MEL202, MEL270, UPMD1, UPMD2, and 92-1 were obtained from Dr. Boris Bastian (PMID: 22236444). KPC cells were obtained from Dr. Ben Stanger (PMID: 21436454). SK-MEL30 cells and RN2 cells were provided by Dr. David B. Solit (PMID: 21725359) and Dr. Scott W. Lowe (PMID: 21131983), respectively. |

| | |
|---|---|
| Authentication | An aliquot of each cell lines were authenticated using ATCC fingerprinting. Otherwise, the cells were submitted for short tandem repeat (STR) profiling and MSK-IMPACT (integration mutation profiling of actionable cancer targets) for mutational status at MSKCC to confirm their authenticity. |
| Mycoplasma contamination | All cell lines are frequently tested for mycoplasma contamination. Cell lines used in this study were verified to be mycoplasma negative before undertaking any experiments with them. |
| Commonly misidentified lines<br>(See ICLAC register) | No commonly misidentified cell lines were used. |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | 6-8 week SCID and NGS mice were purchased from The Jackson Laboratory (stock #001303 and stock #005557, respectively). |
| Wild animals | The study did not involve wild animals. |
| Field-collected samples | The study did not involve samples collected from the field. |
| Ethics oversight | All animals were housed at Memorial Sloan Kettering Cancer Center (MSKCC). All animal procedures were completed in accordance with the Guidelines for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committees at MSKCC. All mouse experiments were performed in accordance with a protocol approved by the MSKCC Institutional Animal Care and Use Committee (11-12-029). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | Patients with chronic lymphocytic leukemia (CLL) and uveal and mucosal melanoma seen at Memorial Sloan Kettering Cancer Center who provided anonymized peripheral blood mononuclear cells (for CLL) and tissue biopsies (for melanoma). |
| Recruitment | Patients with CLL and uveal and mucosal melanoma seen at Memorial Sloan Kettering Cancer Center (MSK) who consented to MSK IRB protocol 06-107 (for CLL) and IRB 14-191 (for melanoma samples) were eligible for inclusion regardless of race, gender, ethnicity or other characteristics. |
| Ethics oversight | Studies were approved by the Institutional Review Boards of Memorial Sloan Kettering Cancer Center (MSK), informed consent was obtained from all subjects (under MSK IRB protocol 06-107) and conducted in accordance to the Declaration of Helsinki protocol. Patients provided samples after their informed consent and primary human de-identified CLL samples derived from whole peripheral blood or BM mononuclear cells were utilized. PDX models were performed using tumor biopsies from de-identified patients under MSK IRB protocol 14-191. Genomic alterations in melanoma tumor biopsies and CLL cells were analyzed using MSK IMPACT28 assay or FoundationOne Heme30 assay, both as previously described. Patient samples were anonymized by the Hematologic Oncology Tissue Bank of MSK (for CLL samples) and the MSK Antitumor Assessment Core Facility (for PDX samples). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## ChIP-seq

### Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| | |
|---|---|
| Data access links<br>*May remain private before publication.* | The data (including related RNA-seq data) has been deposited in the Gene Expression Omnibus under accession number GSE124720 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124720) |
| Files in database submission | The above accession includes related RNA-seq data. The relevant ChIP-seq files (.fastq, .bigwig and associated peak calls) can be obtain at the link above:<br><br>GSM3544057 MEL270.dBRD9.BRG1<br>GSM3544058 MEL270.dBRD9.GLTSCR1<br>GSM3544059 MEL270.dBRD9.input<br>GSM3544061 MEL270.DMSO.BRD9<br>GSM3544062 MEL270.DMSO.BRG1<br>GSM3544063 MEL270.DMSO.GLTSCR1<br>GSM3544064 MEL270.DMSO.input<br>GSM3544066 MEL270.SF3B1_K700E.BRG1<br>GSM3544067 MEL270.SF3B1_K700E.GLTSCR1<br>GSM3544068 MEL270.SF3B1_K700E.input |

| | |
|---|---|
| | GSM3544070 MEL270.SF3B1_WT.BRD9<br>GSM3544071 MEL270.SF3B1_WT.BRG1<br>GSM3544072 MEL270.SF3B1_WT.GLTSCR1<br>GSM3544073 MEL270.SF3B1_WT.input |
| Genome browser session<br>(e.g. UCSC) | https://genome.ucsc.edu/s/chewguoliang/SF3B1%20BRD9%20MEL270%20ChIP |

## Methodology

| | |
|---|---|
| Replicates | The ncBAF complex was pulled-down with two distinct antibodies against BRD9 and GLTSCR1 in two separate control and ncBAF perturbation experiments that used orthogonal methods to perturb ncBAF formation. Analyses used the intersection of data from these distinct experiments. |
| Sequencing depth | Sequencing depth, as well as mapping rates are described below:<br><br>Sample Reads_total Reads_mapped Reads_unmapped Reads_multimapped<br>MEL270.dBRD9.GLTSCR1 21663452 17327309 (79.98%) 3317111 (15.31%) 1019032 (4.70%)<br>MEL270.dBRD9.input 18584217 16915436 (91.02%) 761948 (4.10%) 906833 (4.88%)<br>MEL270.dBRD9.BRG1 20392804 18016161 (88.35%) 1654723 (8.11%) 721920 (3.54%)<br>MEL270.DMSO.BRD9 19785290 15734102 (79.52%) 2934586 (14.83%) 1116602 (5.64%)<br>MEL270.DMSO.GLTSCR1 19281052 15911493 (82.52%) 2501078 (12.97%) 868481 (4.50%)<br>MEL270.DMSO.input 19743934 17876186 (90.54%) 899111 (4.55%) 968637 (4.91%)<br>MEL270.DMSO.BRG1 21373496 18958611 (88.70%) 1645389 (7.70%) 769496 (3.60%)<br>MEL270.SF3B1_K700E.GLTSCR1 20098772 15780486 (78.51%) 3347543 (16.66%) 970743 (4.83%)<br>MEL270.SF3B1_K700E.input 18866831 17097903 (90.62%) 832672 (4.41%) 936256 (4.96%)<br>MEL270.SF3B1_K700E.BRG1 19879373 17670132 (88.89%) 1491006 (7.50%) 718235 (3.61%)<br>MEL270.SF3B1_WT.BRD9 15700266 12196453 (77.68%) 2546295 (16.22%) 957518 (6.10%)<br>MEL270.SF3B1_WT.GLTSCR1 19755923 16270285 (82.36%) 2565754 (12.99%) 919884 (4.66%)<br>MEL270.SF3B1_WT.input 19401125 17523685 (90.32%) 895478 (4.62%) 981962 (5.06%)<br>MEL270.SF3B1_WT.BRG1 19408244 17112467 (88.17%) 1589600 (8.19%) 706177 (3.64%) |
| Antibodies | BRG1 (Abcam EPNCIR111A, Lot # GR3208604-8)<br>GLTSCR1 (Santa Cruz SC-240516, Lot # A2313)<br>BRD9 (Abcam, ab137245) |
| Peak calling parameters | Narrow peaks were called using the callpeak function from MACS2 v2.1.1.20160309 against matched input samples, using default parameters and a p-value cutoff of 1e-5. |
| Data quality | For all samples, a p-value cutoff of 1e-5 against input was used. All peaks were called at a q-value of < 0.017. For each sample, the number of peaks with a fold-change > 5, and the total number of peaks called is as follows:<br>MEL270.dBRD9.BRG1 68229 94156<br>MEL270.dBRD9.GLTSCR1 2982 4504<br>MEL270.DMSO.BRD9 13390 20591<br>MEL270.DMSO.BRG1 73624 101272<br>MEL270.DMSO.GLTSCR1 4453 6564<br>MEL270.SF3B1_K700E.BRG1 60362 85463<br>MEL270.SF3B1_K700E.GLTSCR1 6939 11118<br>MEL270.SF3B1_WT.BRD9 17434 25208<br>MEL270.SF3B1_WT.BRG1 64209 89486<br>MEL270.SF3B1_WT.GLTSCR1 5794 8678 |
| Software | ChIP-seq reads were mapped to the genome by calling Bowtie v1.0.048 with the arguments '-v 2 -k 1 -m 1 --best --strata'. Peaks were called using MACS2 v2.1.1.2016030952 against input control libraries with p < 10-5 and subsequently filtered to remove peaks contained within ENCODE blacklisted regions and the mitochondrial genome. Subsequent data analysis was performed with Bioconductor in the R programming environment. Consensus peaks between samples were called using the soGGI package v1.14.0. Peaks were annotated using the ChIPseeker package v1.18.0. |

# CORRECTIONS & AMENDMENTS

CORRECTION

https://doi.org/10.1038/s41586-019-1602-8

## Author Correction: Increasing $CO_2$ threatens human nutrition

Samuel S. Myers, Antonella Zanobetti, Itai Kloog, Peter Huybers, Andrew D. B. Leakey, Arnold J. Bloom, Eli Carlisle, Lee H. Dietterich, Glenn Fitzgerald, Toshihiro Hasegawa, N. Michele Holbrook, Randall L. Nelson, Michael J. Ottman, Victor Raboy, Hidemitsu Sakai, Karla A. Sartor, Joel Schwartz, Saman Seneweera, Michael Tausz & Yasuhiro Usui

Correction to: *Nature* https://doi.org/10.1038/nature13179, published online 07 May 2014.

In this Letter, we ought to have cited four relevant references. The first two[1,2], published after submission of the first version of our manuscript, report a cultivar-dependent reduction in wheat grain quality under elevated carbon dioxide ($CO_2$) in an Australian Grains Free-Air Carbon dioxide Enrichment (AGFACE) experiment. A subset of these results was reported in two papers published in 2012 (refs. [3,4]). Although our work was independent of these papers, we regret not citing them to highlight their relevance. The original Letter has not been corrected.

1. Fernando, N. et al. Intra-specific variation of wheat grain quality in response to elevated [$CO_2$] at two sowing times under rain-fed and irrigation treatments. *J. Cereal Sci.* **59**, 137–144 (2013).
2. Fernando, N. et al. Elevated $CO_2$ alters grain quality of two bread wheat cultivars grown under different environmental conditions. *Agric. Ecosyst. Environ.* **185**, 24–33 (2013).
3. Fernando, N. et al. Rising atmospheric $CO_2$ concentration affects mineral nutrient and protein concentration of wheat grain. *Food Chem.* **133**, 1307–1311 (2012).
4. Fernando, N. et al. Wheat grain quality under increasing atmospheric $CO_2$ concentrations in a semi-arid cropping system. *J. Cereal Sci.* **56**, 684–690 (2012).

CORRECTION

https://doi.org/10.1038/s41586-019-1628-y

# Author Correction: Human placenta has no microbiome but can contain potential pathogens

Marcus C. de Goffau, Susanne Lager, Ulla Sovio, Francesca Gaccioli, Emma Cook, Sharon J. Peacock, Julian Parkhill, D. Stephen Charnock-Jones & Gordon C. S. Smith

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1451-5, published online 31 July 2019.

In this Article, the Supplementary Information file containing the 16S rRNA sequencing results has been adapted to include a second tab labelled 'ENA link' that provides direct links between individual raw data files in the European Nucleotide Archive (ENA) and the sample numbers used in the Article. This has been added to make it easier for researchers to retrieve 16S rRNA gene datasets from the ENA and link these with the relevant metadata; previously, it was possible to link the ENA files to the corresponding sample via the cram file names, but this was overly complicated and the relationships were not obvious.

In addition, in the Source Data files for Fig. 1 and Extended Data Fig. 5, the arbitrary numbers given to the placental samples did not correspond to the same samples in cohorts 1 and 2. Matching sample numbers have now been provided in both Source Data files, and have also been added to the Supplementary Information sequencing file. Two minor errors in the headings in the Source Data file for Extended Data Fig. 5 (cells F3 and AM1) have also been corrected. We also noticed that the wrong file had been provided as Source Data for Fig. 3 and this has now been replaced. For transparency to readers, the original Supplementary Information sequencing file and the original Source Data files for Figs. 1, 3 and Extended Data Fig. 5 are provided as Supplementary Information to this Amendment.

Finally, a minor error in the name of the tab in the Source Data file for Extended Data Fig. 7 (originally named 'Extended Data Fig. 5') has also been corrected.

All of these additions and corrections do not change any of the data or conclusions reported in the Article, and are only necessary to facilitate reanalysis of our data by other researchers.

**Supplementary information** is available in the online version of this Amendment.

# CORRECTIONS & AMENDMENTS

**CORRECTION**

https://doi.org/10.1038/s41586-019-1626-0

## Publisher Correction: Natural selection on the *Arabidopsis thaliana* genome in present and future climates

Moises Exposito-Alonso, 500 Genomes Field Experiment Team, Hernán A. Burbano, Oliver Bossdorf, Rasmus Nielsen & Detlef Weigel

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1520-9, published online 28 August 2019.

In this Letter, owing to an error in the production process, author Moises Exposito-Alonso was inadvertently associated with the affiliation 'Globe Institute, University of Copenhagen, Copenhagen, Denmark', and instead should have been associated with the present address: 'Department of Integrative Biology, University of California Berkeley, Berkeley, CA, USA.'. The Letter has been corrected online.

CORRECTION

https://doi.org/10.1038/s41586-019-1627-z

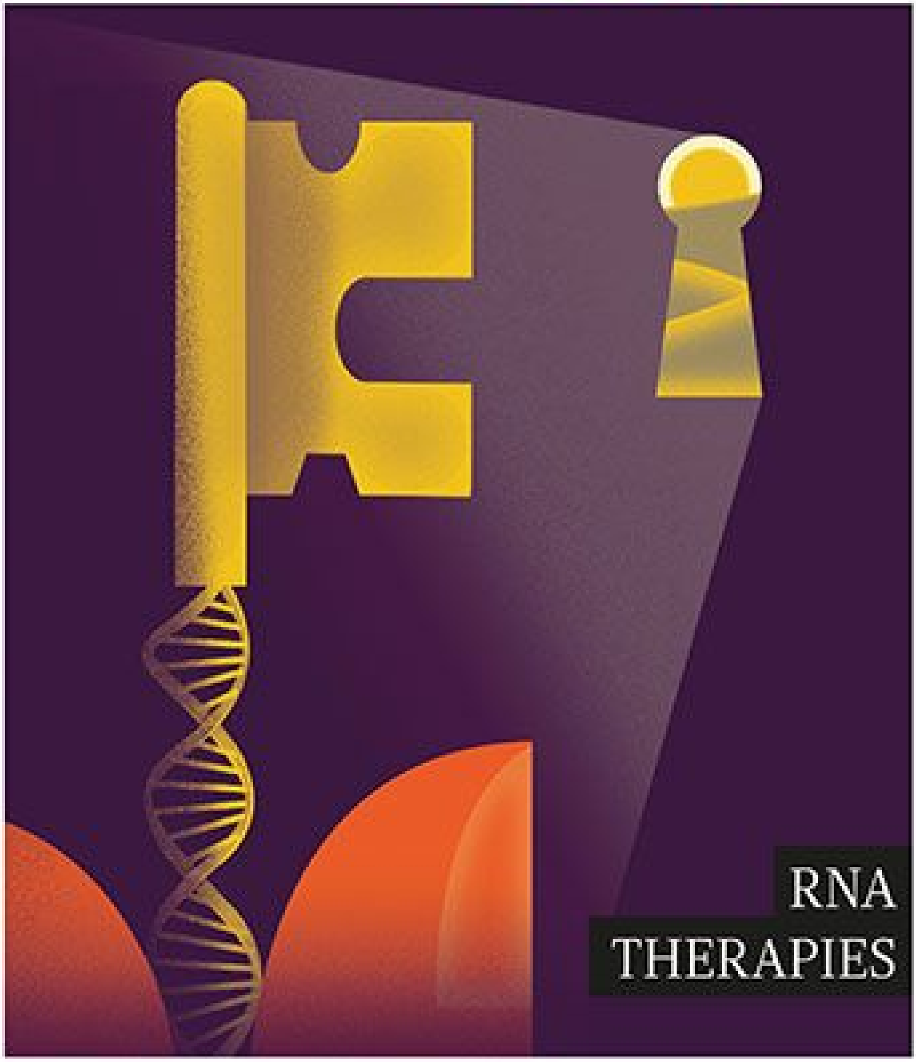
# Publisher Correction: An actin-based viscoplastic lock ensures progressive body-axis elongation

Alicia Lardennois, Gabriella Pásti, Teresa Ferraro, Flora Llense, Pierre Mahou, Julien Pontabry, David Rodriguez, Samantha Kim, Shoichiro Ono, Emmanuel Beaurepaire, Christelle Gally & Michel Labouesse

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1509-4, published online 28 August 2019.

In Figs. 1c and 3a of this Letter, the genotype should have read '*Pdpy-7::LifeAct::GFP*' instead of '*Ppdy-7::LifeAct::GFP*'. In addition, the black box in Extended Data Fig. 1d should have shown the text '*spc-1*(RNAi) *pix-1*(*gk416*)'. These errors have been corrected online.

natureOUTLOOK



# RNA THERAPIES

## Unlocking the molecule's potential

Produced with support from:

IONIS

# natureOUTLOOK

# RNA THERAPIES

17 October 2019 / Vol 574 / Issue No 7778

Cover art: Neil Webb

**Editorial**
Herb Brody, Richard Hodson, Elizabeth Batty, Jenny Rooke, Nick Haines
**Art & Design**
Mohamed Ashour, Andrea Duffy, Denis Mallet
**Production**
Nick Bruni, Karl Smart, Ian Pope, Kay Lewis
**Sponsorship**
David Bagshaw, Claudia Danci
**Marketing**
Nicole Jackson
**Project Manager**
Rebecca Jones
**Creative Director**
Wojtek Urbanek
**Publisher**
Richard Hughes
**VP, Editorial**
Stephen Pincock
**Managing Editor**
David Payne
**Magazine Editor**
Helen Pearson
**Editor-in-Chief**
Magdalena Skipper

Carlos Heras-Palou was a 39-year-old surgeon with two young daughters when he developed hereditary transthyretin (hATTR) amyloidosis, a rare degenerative disease caused by the misfolding of a protein called transthyretin (TTR). That diagnosis in 2004 carried a grim prognosis: a typical life expectancy of 3 to 15 years. At the time, most people who developed hATTR amyloidosis needed to be treated with transplant surgery. But another option emerged. Heras-Palou and his younger sister Isabel, who was diagnosed with the condition in 2013, enrolled in studies for a drug that thwarts production of the toxic protein by binding to messenger RNA — the molecule that carries genetic instructions for producing TTR.

Approval of this drug and one other by US and European regulators in 2018 (see page S7) has given fresh impetus to the nascent research field after more than a decade of false starts. That development has raised hopes that RNA therapies in their various forms (S2) can treat not only rare conditions (S16) but also more common ones, including treatments for cardiovascular disease (S13) and vaccines for cancer (S10).

More than a dozen RNA therapies are currently being tested in clinical trials. But major challenges remain, most notably in how to deliver therapeutic strands of RNA into the right cells (S8). Progress will depend on a continuing commitment from pharmaceutical companies to drive and fund innovation (S4). Researchers also need to up their game when it comes to communicating the underlying science of gene manipulation in all its complexity to patients and the wider public (S15).

We are pleased to acknowledge the financial support of Ionis Pharmaceuticals in producing this Outlook. As always, *Nature* retains sole responsibility for all editorial content.

**David Payne**
*Managing Editor, Nature Careers and Supplements*

## CONTENTS

*Nature Outlooks* are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the Nature Outlook Editorial guidelines available at go.nature.com/e4dwzw

**CITING THE OUTLOOK**
Cite as a supplement to *Nature*, for example, *Nature* **Vol. XXX**, No. XXXX Suppl., Sxx–Sxx (2019).

**VISIT THE OUTLOOK ONLINE**
The *Nature Outlook RNA therapies* supplement can be found at www.nature.com/collections/rna-therapy-outlook
It features all newly commissioned content as well as a selection of relevant previously published material that is made freely available for 6 months.

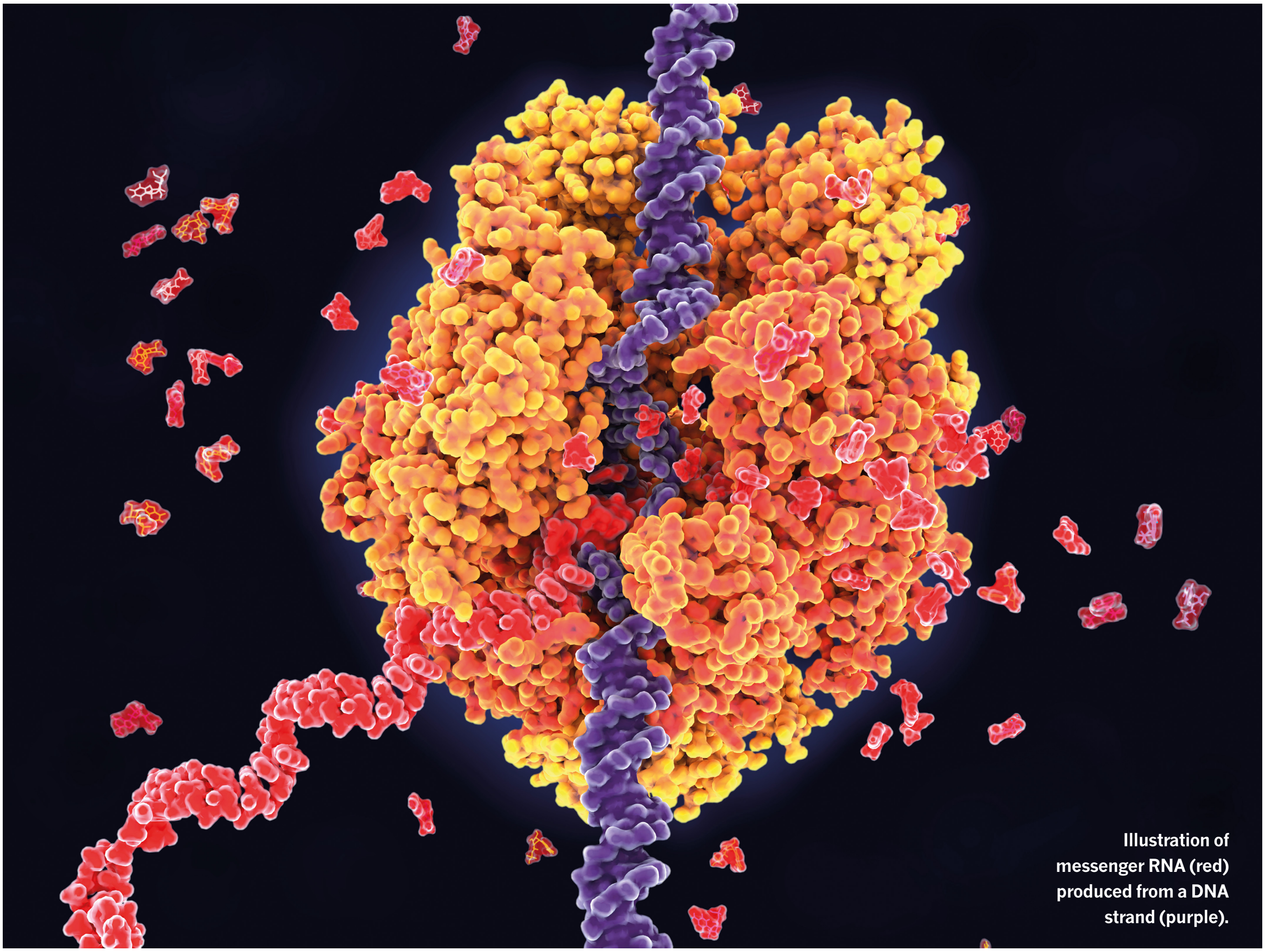
**SUBSCRIPTIONS AND CUSTOMER SERVICES**
Site licences (www.nature.com/libraries/site_licences): Americas, institutions@natureny.com; Asia-Pacific, http://nature.asia/jp-contact; Australia/New Zealand, nature@macmillan.com.au; Europe/ROW, institutions@nature.com; India, npgindia@nature.com. Personal subscriptions: UK/Europe/ROW, subscriptions@nature.com; USA/Canada/Latin America, subscriptions@us.nature.com; Japan, http://nature.asia/jp-contact; China, http://nature.asia/china-subscribe; Korea, www.natureasia.com/ko-kr/subscribe.

**CUSTOMER SERVICES**
Feedback@nature.com

# RNA THERAPIES EXPLAINED

*Treatments that target RNA or deliver it to cells fall into three broad categories, with hybrid approaches also emerging.*

**BY SARAH DEWEERDT**



JUAN GAERTNER/SPL

**Illustration of messenger RNA (red) produced from a DNA strand (purple).**

## PATH TO THE CLINIC

*Messenger RNA was discovered in 1961, but RNA therapies only took off in the 1990s.*

| 1990 | 1993 | 1998 | 1998 | 2001 | 2002 |
|---|---|---|---|---|---|
| A study in mice shows that injecting mRNA into skeletal muscle leads to production of the protein encoded by the RNA[1]. This result lays the groundwork for treatments based on mRNA. | Injection of mRNA from the influenza virus induces an immune response in mice[2], providing a proof of concept for RNA-based vaccines. | A team led by Andrew Fire at the Carnegie Institution for Science, Washington DC, and Craig Mello at the University of Massachusetts, Worcester, show that short interfering RNAs (siRNA) can suppress gene activity in *Caenorhabditis elegans*[3]. | The first RNA therapy, fomivirsen, is approved by the US Food and Drug Administration (FDA). The antisense oligonucleotide (ASO) drug tackles cytomegalovirus retinitis, inflammation of the retina. | Researchers demonstrate that RNA interference (RNAi), a mechanism for gene silencing underpinned by siRNA, occurs not only in plants and invertebrates, but also in mammalian cells[4], suggesting its potential for targeting harmful genes. | Researchers led by Mark Kay at Stanford University, California, use RNAi to target a sequence in the hepatitis C virus for destruction in mice[5], highlighting the technology's therapeutic potential. |

The conventional view of RNA casts the molecule in a supporting role — the intermediary between DNA and protein, and a passive conduit for information.

And that's how the most familiar form of RNA, messenger RNA, works. But only a small fraction of RNA molecules in cells are mRNAs.

As well as carrying instructions for making proteins, RNAs help to turn genes on and off, aid chemical reactions, slice and dice other RNAs, and even build proteins — by transporting amino acids and linking them together.

"They're much more active players in cell biology than we initially thought," says Bruce Sullenger, who investigates RNA therapies at Duke University School of Medicine in Durham, North Carolina.

In turn, these diverse roles have inspired a host of ideas about how to harness RNA for use in medicine. These developments culminated in the 2018 approval, in both the United States and Europe, of two RNA-based therapies for hereditary ATTR amyloidosis — a progressive and potentially fatal disorder in which abnormal proteins build up in nerves and organs such as the heart. Many more RNA therapies are in the development pipeline, with around a dozen already being tested in clinical trials.

Most RNA therapies can be sorted into one of three broad categories: those that target nucleic acids (either DNA or RNA), those that target proteins, and those that encode proteins. Hybrid approaches that combine several RNA-based mechanisms into a single package are also emerging.

There are two distinct types of therapies that target nucleic acids: single-stranded antisense oligonucleotides (ASOs), and double-stranded molecules that operate through a cellular pathway known as RNA interference (RNAi).

ASOs are short stretches of modified DNA made up of about 13–25 building blocks, or nucleotides. These molecules prevent mRNA from being translated into protein by several mechanisms, including blocking the start of translation or tagging the mRNA for degradation. One of the amyloidosis drugs approved last year, inotersen, is an ASO.

ASOs can also alter splicing, the process that sculpts a precursor messenger RNA into its mature form. Two of this type of ASO received FDA approval in 2016: nusinersen, which targets a fatal inherited condition called spinal muscular atrophy, and eteplirsen, a treatment for Duchenne muscular dystrophy. The latter is an example of an 'exon' skipping drug, which uses an ASO to block only the mutated portion of a gene from being expressed. The result is a protein that is functional, but that lacks the mutated portion that causes a pathology (see page S16).

Because RNAi makes use of double-stranded molecules, these therapies are tougher to get into cells than ASOs. But fewer molecules are needed for the therapy to be effective. RNAi involves small interfering RNAs (siRNAs), 21–23 nucleotides long, or similar molecules such as microRNAs, to degrade mRNA and prevent it from being translated into protein. The other amyloidosis drug approved in 2018, patisiran, is an siRNA therapy.


MARIO TAMA/GETTY IMAGES

A mother and her baby who has microcephaly, which is linked to Zika virus.

RNA therapies that target proteins use a type of molecule known as an RNA aptamer. The molecule is designed to bind to a specific site on a specific protein to modulate its function. Pegaptanib, a treatment for a form of age-related macular degeneration in which blood vessels penetrate the retina and cause vision to deteriorate, is an example of such a drug. Pegaptanib binds to and blocks the function of the protein vascular endothelial growth factor, leading to a reduction in the growth and permeability of blood vessels in the eye. RNA aptamers might be useful in surgery and emergency medicine, in which their rapid action and reversibility could aid anaesthesia and modulate blood clotting.

RNA therapies that use mRNAs are being used to develop personalized cancer vaccines (S10), as well as vaccines for infectious diseases such as Zika virus, which has been linked with the condition microcephaly. Researchers are also exploring whether these type of treatments can be used as protein-replacement therapies for rare conditions such as the blood-clotting disorder haemophilia.

Some researchers are pursuing hybrid strategies that draw on the three main forms of RNA therapy. Apic Bio in Cambridge, Massachusetts, uses an approach that it calls silence and replace, which uses an RNAi component to silence a harmful gene and an mRNA component that encodes a normal version of the corresponding mutated protein. Both are delivered by a single viral vector. The company is doing preclinical work on hybrid drugs aimed at α1-antitrypsin deficiency, an inherited lung and liver disease, and hereditary amyotrophic lateral sclerosis, a degenerative neurological condition.

The biggest barrier to RNA therapy has long been delivering RNA to the correct place in the correct cells. The past several years have seen a flurry of advances that have improved researchers' ability to get such drugs into liver cells — an important development because so many proteins implicated in diseases are made in the liver. ■

---

**Sarah DeWeerdt** *is a science journalist in Seattle, Washington.*

**2003**

Judith Lieberman and her colleagues at Harvard Medical School in Boston, Massachusetts, demonstrate that RNAi can suppress HIV replication in macrophages[6].

**2004**

Pegaptanib becomes the first RNA aptamer that targets proteins to be approved by the FDA.

**2006**

Fire and Mello receive the Nobel Prize in Physiology or Medicine for their work on RNAi. Meanwhile, Kay and his team report that long-term siRNA expression can cause liver damage and even death in mice[7], temporarily halting progress on RNAi-based therapies.

**2010**

The first use of RNAi-mediated gene silencing in humans is reported in a phase I trial in people with the skin cancer melanoma[8]. The therapy reduced expression of a gene needed for tumour cells to multiply.

**2018**

Approval to market patisiran and inotersen for hereditary ATTR amyloidosis is granted in the United States and Europe.

1. Wolff, J. A. *et al. Science* **247**, 1465–1468 (1990).
2. Martinon, F. *et al. Eur. J. Immunol.* **23**, 1719–1722 (1993).
3. Fire, A. *et al. Nature* **391**, 806–811 (1998).
4. Elbashir, S. M. *et al. Nature* **411**, 494–498 (2001).
5. McCaffrey, A. P. *et al. Nature* **418**, 38–39 (2002).
6. Song, E. *et al. J. Virol.* **77**, 7174–7181 (2003).
7. Grimm, D. *et al. Nature* **441**, 537–541 (2006).
8. Davis, M. E. *et al. Nature* **464**, 1067–1070 (2010).


NEIL WEBB

DRUG DISCOVERY

# Pharma’s RNA roller coaster

*After a groundswell of hype and a sceptical backlash, the pharmaceutical industry is learning how to leverage RNA interference in the clinic.*

**BY MICHAEL EISENSTEIN**

Fortunes can shift precipitously in the drug-discovery world. At the start of the twenty-first century, all eyes were focused on a powerful gene-silencing technology called RNA interference (RNAi), and many companies saw the almost limitless potential of harnessing the tool to manipulate genes implicated in diseases. By 2006, Andrew Fire and Craig Mello, the biologists who discovered RNAi, had received the Nobel Prize in Physiology or Medicine, and large pharmaceutical companies were pouring billions of dollars into RNAi start-ups.

Just four years later, this exuberance had given way to despair — and the money was drying up. “People started giving up hope,” says John Maraganore, chief executive of Alnylam Pharmaceuticals — one of the first biotechnology companies to pursue RNAi therapy. “They started thinking that drugs would never come out of this, and the pharmaceutical industry left the space.” The problem was not the mechanism behind RNAi, which quickly became a central element of the genetics-research toolbox, but rather the difficulties associated with safely delivering intact molecules of gene-modulating RNA to target tissues.

Even as investors fled, a small number of companies soldiered on, shifting strategies and therapeutic targets, and finally notching up some victories in clinical trials. In 2018, Alnylam, which is based in Cambridge, Massachusetts, netted the first approval for an RNAi therapy in both the United States and Europe. That drug, patisiran (Onpattro), is used to treat hereditary transthyretin (hATTR) amyloidosis, a build-up of the protein amyloid in nerves and organs. Several other biotech companies have also got RNAi candidate drugs in their pipelines, and the despair from a decade ago is all but forgotten. “There’s now a sense of destiny in the field,” says Douglas Fambrough, chief executive and co-founder of Dicerna Pharmaceuticals in Cambridge. Such confidence was made possible only after learning some costly and difficult lessons — a body of knowledge that has helped to distinguish the true potential of RNAi therapy from the hype that accompanied its birth.

## BILLION-DOLLAR BABY

RNAi-therapy researchers know that many existing medicines received approval only after the technologies that underpin them took a serious tumble. Gene therapy, for example, has a troubled history, in which the death of a teenage boy during a clinical trial in 1999 derailed the field for years; the first such product to receive approval from the US Food and Drug Administration (FDA), voretigene neparvovec (Luxturna), developed by Spark Therapeutics in Philadelphia, Pennsylvania, to treat a rare form of vision impairment, did not receive approval until 2017. Even monoclonal antibodies struggled to receive regulatory approval, with the first generation of mouse-derived molecules generating unwanted immune responses in people. Only after the development of ‘humanized’ monoclonal antibodies did the technology take off in the clinic. “There were multiple moments where people gave up hope,” says Maraganore. “But now monoclonals are the largest class of pharmaceutical medicines.”

The instant appeal of RNAi therapy is understandable. In 1998, Fire, then at the Carnegie Institution for Science in Washington DC, and Mello, at the University of Massachusetts in Worcester, demonstrated that they could efficiently and selectively dial down the expression of various genes in the worm *Caenorhabditis elegans* by injecting small quantities of short

ARNE DEDERT/DPA/PA IMAGES

interfering RNA (siRNA) molecules[1], which comprise paired strands of RNA. Over the next several years, researchers began to understand the molecular mechanisms that underlie RNAi and the design principles for generating siRNAs that inhibit gene expression effectively. After a team led by gene-therapy researcher Mark Kay at Stanford University in California demonstrated[2] the first use of RNAi in mice in 2002, the clinical potential of this method was immediately apparent.

By May 2003, US business magazine *Fortune* was calling RNAi "biotech's billion-dollar breakthrough", and a handful of companies had staked their worth on the technology, including Alnylam, founded the previous year by several RNA biologists, and Sirna Therapeutics, launched in San Francisco, California, in 2003 by Fambrough and other investors. Although there were further contenders, these two companies controlled a formidable array of intellectual property that pertained to the clinical application of RNAi, and big pharma was eager to write them cheques to get a place at the table. In 2006, the same year in which Mello and Fire received the Nobel prize, drug company Merck in Kenilworth, New Jersey, bought Sirna Therapeutics for more than US$1 billion in cash, and Alnylam lined up a formidable array of partnerships with larger firms. "They did these big deals with Takeda, Roche and Novartis," says Ted Tenthoff, managing director at investment bank Piper Jaffray in New York City. "The value of these deals was probably on the order of billions of dollars."

## CRASH ON DELIVERY

Despite the unbridled enthusiasm of investors, early efforts to move RNAi therapies through clinical trials demonstrated that important challenges remained. "When we were doing our initial studies, the big question was what would be the time frame in which one could develop a clinically relevant delivery approach that would work safely in humans," recalls Kay. The most pressing issue was working out how to get siRNAs into the correct cells in the body, and in an intact state. "The delivery technology just wasn't there," says Ritu Baral, managing director at Cowen, an investment bank in New York City. "You would give somebody RNAi and their enzymes would just rip the thing to shreds immediately, before it got near where it needed to go."

Another problem was that these initial RNAi therapies were often mistaken by the body's immune system as the remnants of infectious agents, which would trigger side effects. "Experiments in animal models miss the fact that the human immune response is much more sensitive," says Maraganore. "We had to learn more about how these molecules behaved in a human setting." The first clinical programmes, which targeted infectious diseases and eye disorders, yielded little progress — a stagnation that made RNAi therapy's early backers increasingly nervous. In the late 2000s, that concern reached a crisis point.

Baral puts it bluntly: "Big pharma lost its nerve." In 2010, Roche and Novartis, both in Basel, Switzerland, began to pull back, terminating their partnerships with Alnylam. The following year, Pfizer in New York City and Abbott in Abbott Park, Illinois, ended their independent RNAi drug-discovery programmes. Merck scaled back its RNAi efforts, but held out for a few more years before selling its Sirna Therapeutics assets to Alnylam in 2014 for less than 20% of the price it had paid in 2006. "The progress towards clinical-stage programmes wasn't as fast as the field had collectively thought it was going to be," says James Hamilton, vice-president of clinical development at Arrowhead Pharmaceuticals, an RNAi company in Pasadena, California. More importantly, the bad news had a ripple effect on media coverage of the industry, creating a narrative in which RNAi had failed broadly as a clinical tool.

The RNAi companies that survived this time did so through a combination of determination and sacrifice. Alnylam maintained a financial lifeline through a few committed investors and by licensing its technologies, but it also had to make some tough decisions. "We had to do two rounds of lay-offs around that time to preserve enough capital to get the data that we ultimately needed to convince investors," says Maraganore. The company's competitors, in the meantime, made double-or-nothing bets. At Arrowhead, for example, Hamilton and his colleagues took advantage of big pharma's flight by convincing investors to back acquisition of the RNAi portfolios being ditched by other companies — from Roche in 2011 and Novartis in 2015. "We were a new company focused purely on RNAi after the Roche transaction, whereas it was just a piece of what we were doing before," says Hamilton.

> ***"We always had good reason to believe that we could make it work."***

## BACK ON TRACK

The irony is that the exodus occurred just as clinical RNAi efforts were starting to bear fruit. Long-standing researchers remained confident, even as scepticism swelled. "We always had good reason to believe that we could make it work," says Fambrough, who launched Dicerna in 2007, shortly after Merck purchased Sirna Therapeutics. The clinical-trials data were finally beginning to justify this confidence. A 2010 study clearly demonstrated that siRNA could effectively silence specific genes in humans[3], and by 2013, Alnylam had published clinical-trials data that demonstrated patisiran's safety and efficacy[4].

**Biologists Craig Mello (left) and Andrew Fire discovered RNA interference.**

By this time, researchers at such companies had learnt important lessons about working with RNAi technology in people. For example, it became clear that almost all siRNA-based drugs administered into the bloodstream tend to accumulate in the liver. RNAi companies therefore turned their focus to conditions that could be treated by targeting this organ, such as hereditary transthyretin amyloidosis. It also became clear that making chemical modifications to RNA can greatly improve its performance as a drug. "They enhance the stability of the molecule, and reduce the immunogenicity," explains Kay, who notes that other types of modification can improve the efficiency with which therapeutic RNA molecules are incorporated into the cellular machinery that underpins RNAi. "We didn't understand a lot of stuff back in 2001 or 2002."

Robust solutions to the problem of delivering RNAi therapies to target tissues have also emerged. Patisiran is encapsulated by lipid nanoparticles, an approach that was sufficiently safe and effective for delivery to the liver to win regulatory approval. But Alnylam and other companies have shifted to an alternative delivery strategy, in which siRNAs are chemically coupled to the sugar *N*-acetylgalactosamine (GalNAc), which binds strongly to a receptor that is expressed abundantly in the liver. "GalNAc delivery to the liver is better than lipid nanoparticles in every way," says Fambrough. "There's less toxicological burden, it's easier to make, it lasts longer and it's easier to administer." Ease of administration is a crucial advantage: GalNAc enables the subcutaneous injection of RNAi therapies, rather than requiring their intravenous delivery, as do lipid-based formulations.

Even with these advances, the past five years have seen major setbacks.

Alnylam's first experiment with GalNAc-mediated RNAi delivery, revusiran, a treatment for hereditary transthyretin amyloidosis, ended abruptly in 2016 after a phase III trial in people with amyloidosis-associated heart disease showed that there were more deaths in people who received the treatment than

A researcher in the lab at Alnylam — the first company to get approval for an RNA-interference therapy.

ALNYLAM PHARMACEUTICALS

in those who took a placebo. The underlying cause of the deaths, which were mostly heart-related, remains unclear, but revusiran's formulation seems to have contributed. "It required very high dose levels to achieve reduced gene expression," says Maraganore. "The data seem to suggest that metabolites of that drug given at those doses might have been poorly tolerated in that very frail population."

Safety concerns also led Arrowhead to take a dramatic decision in 2016, when it pulled the plug on a trio of clinical programmes on the basis of a tentative link between their RNAi delivery strategy and excessive deaths in non-human primates. "It was a tough business decision at the time, but I think, in retrospect, it was the right decision," says Hamilton. He notes that the company has since turned to a GalNAc-based strategy that enables more efficient and durable reductions in gene expression, with no safety concerns emerging from studies in animals and people.

## A BETTER TOMORROW

In 2018, patisiran broke away from the pack to win regulatory approval on the strength of a pivotal trial that showed it had a good safety profile and led to meaningful improvement in the health and quality of life of people with hereditary transthyretin amyloidosis. The decision vindicated researchers who had stuck by RNAi, and particularly those at Alnylam. "It was a moment like no other — once in a lifetime — when you've been involved in 16 years' worth of effort to bring an important innovation forward for patients, and finally have it approved," says Maraganore.

Rival firms such as Arrowhead and Dicerna also benefited from having a first-in-class drug on the market. "This is the first safety package the FDA has thoroughly vetted and passed," says Baral. "And it shows investors that there are markets that one can go after with this technology where there is a likely return on investment." It should also be noted that, despite its proven safety and efficacy, patisiran employs the relatively outdated lipid-nanoparticle delivery system, which requires intravenous delivery and steroids to manage the side effects from infusion. This could give RNAi therapies that use GalNAc-based delivery, which can be achieved with a minimally invasive subcutaneous injection, a chance to shine even more brightly. Indeed, Alnylam is conducting phase III trials of vutrisiran, a GalNAc-modified version of patisiran that can be administered subcutaneously at a low dose every three months, rather than every three weeks.

> ***"It was a moment like no other — once in a lifetime."***

Several RNAi programmes for liver conditions are also under way, including the development of Alnylam's givosiran, which helps to prevent the production of neurotoxic metabolites of haem, a molecule found in red blood cells, that accumulate in people with acute hepatic porphyria. In a 2019 trial, givosiran reduced the frequency of the painful and debilitating neurological symptoms of porphyria attacks by 74%, and the drug is awaiting regulatory approval. Dicerna is performing phase II testing of DCR-PHXC, an siRNA-based drug for primary hyperoxaluria. In this condition, the liver generates excessive levels of a compound called oxalate, which accumulates in the kidneys and can lead to organ failure. And since retooling its delivery strategy, Arrowhead has moved ARO-AAT, a potential treatment for α1-antitrypsin (AAT) deficiency, into a pivotal phase II/III trial. This lung and liver disorder arises from a mutation that causes the AAT enzyme to misfold and then aggregate in cells, and unpublished phase-I data have demonstrated the capacity of ARO-AAT to efficiently decrease production of the mutated protein. "Even at low doses, we were seeing mean serum reductions of 60–70%," says Hamilton.

Buoyed by these successes, the RNAi field is looking beyond the liver. There is particular interest in hitting a tricky target: the central nervous system (CNS). Although getting RNA across the blood–brain barrier remains impractical, progress towards another class of RNA-based therapy has demonstrated the feasibility of repairing neurological damage through direct injections of such drugs into the cerebrospinal fluid. The strategy requires the targeted delivery of single strands of RNA, known as antisense oligonucleotides, to affected tissues. But, in contrast to siRNAs, which undergo processing by enzymes in cells and then modulate gene expression as part of a larger inhibitory protein complex, antisense oligonucleotides inhibit gene expression by directly binding to target messenger RNAs.

Both Alnylam and Dicerna are aggressively pursuing RNAi drugs for conditions that affect the CNS, including Huntington's disease and Alzheimer's disease, and Fambrough is optimistic about the prospects of this line of research. Unlike therapies that target the liver, which rely on a receptor expressed by the organ to access liver cells, he notes that therapeutic RNAs can penetrate neurons without the need for GalNAc modification once they have been delivered into the CNS.

These ambitious efforts are being made possible by fresh investment from big pharma — perhaps the surest sign that RNAi therapy has regained its stride. In October 2018, for example, Eli Lilly in Indianapolis, Indiana, proffered an upfront investment of $100 million to support Dicerna's work in CNS and other non-liver conditions, and Johnson & Johnson in New Brunswick, New Jersey, purchased the rights to Arrowhead's hepatitis B RNAi programme in a deal that could net the biotechnology company up to $1.6 billion. And in April, Regeneron Pharmaceuticals in Tarrytown, New York, and Alnylam embarked on a $1-billion collaboration to develop RNAi therapies for use in the CNS, eye and liver. After almost two decades of development, and with one drug across the regulatory finish line — and several other candidates close behind — hopes are high that these investments might finally unlock RNAi's clinical potential. "I think we've pretty much put the old scepticism and demons to bed," says Fambrough. ■

---

**Michael Eisenstein** *is a freelance journalist based in Philadelphia, Pennsylvania.*

1. Fire, A. *et al. Nature* **391**, 806–811 (1998).
2. McCaffrey, A. P. *et al. Nature* **418**, 38–39 (2002).
3. Davis, M. E. *et al. Nature* **464**, 1067–1070 (2010).
4. Coehlo, T. *et al. N. Engl. J. Med.* **369**, 819–829 (2013).

PERSPECTIVE


UHDB TRUST

# Patisiran's path to approval

**Carlos Heras-Palou** explains his part in making patisiran the first RNA-interference therapeutic available to people with rare diseases.

In 2004, after a few months of experiencing increasing pain in my hands and feet, I was diagnosed with hereditary transthyretin (hATTR) amyloidosis, a progressive degenerative disease. Following diagnosis, most people have a life expectancy of 3 to 15 years.

This rare condition, which affects about 10,000 people worldwide, is caused by the misfolding of a protein called transthyretin (TTR) that is produced in the liver and which normally carries both the thyroid hormone thyroxine and vitamin A in the blood.

At the time, I was 39, married with two children aged 1 and 2, and had my dream job working as a surgeon. After my diagnosis, I could not see how we would manage as a family if I had to stop working because the disease was affecting my hands. I was worried about telling my brother and sister, and the impact of the news on them and their families. The prognosis was very bad, but after a short period of despair, I felt I had to put up a fight even if I could see no chance of winning it.

I went to see Philip Hawkins, who was then clinical director at the National Amyloidosis Centre in London. At the time of my diagnosis, two treatment strategies seemed logical. The first was to find a molecule to stabilize TTR, prevent misfolding and stop abnormal proteins from forming. The second was to silence the gene that encodes TTR, to block production and halt the progression of the disease.

> KNOWING THAT A DRUG EXISTS THAT CAN STOP THE **PROGRESSION** OF A DISEASE YET IS NOT AVAILABLE TO PATIENTS IS **FRUSTRATING.**

A liver transplant was one option to slow disease progression, because a transplanted organ doesn't produce the abnormal protein. Blocking production of TTR in the liver could deliver similar results. At the time, regulating or editing a gene was nowhere near possible. I didn't think anyone would invest a fortune to treat such a rare disease.

Researchers at Alnylam, a drug company in Cambridge, Massachusetts, founded in 2002 to develop RNA interference (RNAi) therapeutics, had mastered the technique of producing snippets of double-stranded RNA known as small interfering RNA (siRNA). They began to look for a disease model to test their laboratory discovery, ideally one with an unmet need, caused by a known protein, with a known, validated gene target, made in an organ to which a drug could be delivered, and for which gene-silencing would result in the reduction of a measurable biomarker. After discussion between the Alnylam research team and Hawkins, hATTR amyloidosis was chosen.

Within a year, the company was reliably manufacturing siRNAs. But these early molecules were plagued by a surplus of negative charges, making them prone to degradation. A delivery system had to be designed and tested — a task that proved more difficult than initially expected, taking 10 years in total. In a clinical trial, ALN-TTR01 was tested on a group of 32 people, but the drug lowered the level of TTR in the blood in only one person. Then a second carrier for siRNA was tried and seemed to work well — *in vitro* in a transgenic mouse model and in a phase I human study.

In 2013, my younger sister was also diagnosed with hATTR amyloidosis. She was offered a liver transplant or the option to enrol in the phase II trial of Alnylam's new drug: ALN-TTR02, also known as patisiran.

She was one of 29 people who received different doses of the drug. The results seemed spectacularly good for patients on the higher doses.

In the subsequent phase III trial, two-thirds of participants received patisiran and one-third received a placebo. I was one of the 225 people enrolled in that study, which ran from November 2013 until August 2017. The results were published in July 2018 (D. Adams *et al. N. Engl. J. Med.* **379**, 11–21; 2018) and found that the drug reduced TTR production by about 81%. The following month, patisiran was approved by both the US Food and Drug Administration and the European Medicines Agency (EMA). I was the patient representative at the EMA's meeting in London. It felt like a historic occasion. After 16 years of work and an investment of about US$3.5 billion by Alnylam into RNAi therapeutics, this was the start of a new era for people with hATTR. I was also a patient representative on the UK National Institute for Health and Care Excellence committee that decided in August to recommend that patisiran be used to treat damage to peripheral nerves due to hATTR amyloidosis.

Knowing that a drug exists that can stop the progression of a disease yet is not available to patients is extremely frustrating. Watching people get worse and die while a drug goes through the regulatory process is heartbreaking. But from a scientific point of view, it has been only 20 years since a new biological observation resulted in a drug being made widely available to patients. Twenty years is short from a scientific-development point of view. For a patient, it feels like a lifetime.

Hereditary ATTR amyloidosis is an autosomal dominant condition, which means there is a 50% chance of each of my children having the gene for the disease. The proportion of people who carry the mutation and show symptoms is variable for unknown reasons. In Sweden, for example, it is around 20%, and in Portugal around 80%.

But from my perspective, the future looks bright for people with hATTR amyloidosis and their families. We now know that there is an effective treatment and that our children will not have to go through the pain and anxiety of this devastating disease.

A new form of the drug under development can be injected subcutaneously once every three months, instead of the current intravenous infusion every three weeks. This is thanks to advances in the chemistry of stabilizing the RNA. In the future, many diseases currently without treatment will be easily managed by a yearly injection of a specific siRNA.

Fifteen years after my initial diagnosis I remain reasonably well, with a good quality of life. I continue to work as a hand and wrist surgeon and thoroughly enjoy my job, carrying out operations and teaching the next generation of hand surgeons. I feel fortunate to be one of the first people to benefit from this new class of drug.

There are 7,000 recognized rare diseases, and hundreds of them should be treatable with RNAi therapeutics. Everything is out there to be done, and everything seems possible. ■

---

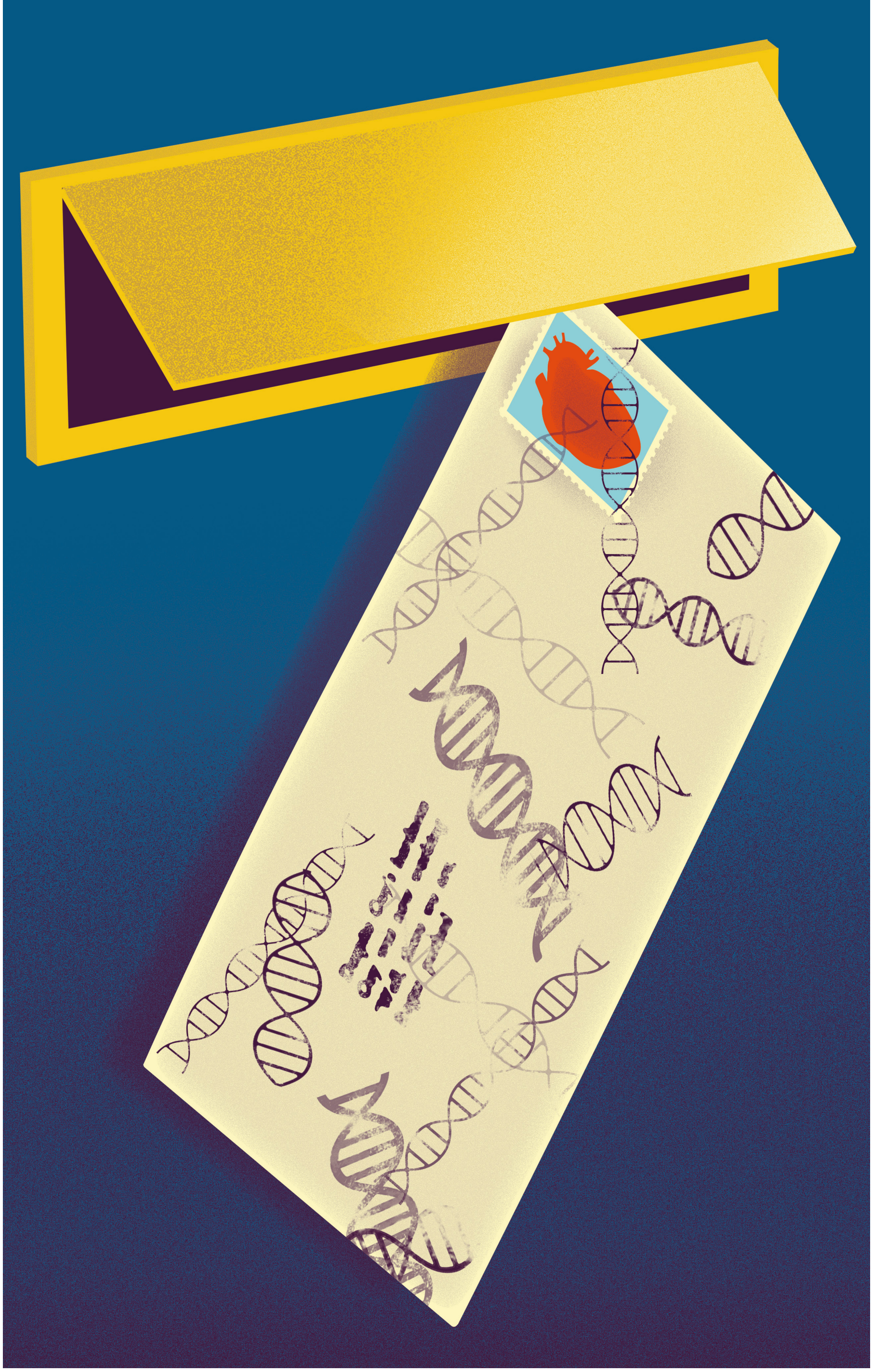
**Carlos Heras-Palou** *is a hand and wrist surgeon at University Hospitals of Derby and Burton NHS Foundation Trust.*
*e-mail: carlos.heraspalou2@gmail.com*

RNA INTERFERENCE

# Strand and deliver

*The difficulties of delivering RNA interference therapeutics to their target cells brought the nascent research field to its knees, but a landmark drug approval offers hope.*

**BY BIANCA NOGRADY**



NEIL WEBB

The commercial story of RNA interference (RNAi) harbours more plot twists and unexpected demises than television fantasy drama *Game of Thrones*.

As in all good dramas, there was someone — or something — that, just as things were looking promising, showed up to foil everyone's plans, in this case hindering efforts to safely get the RNAi into the right cells to go about its work.

Christopher Anzalone, chief executive of Arrowhead Pharmaceuticals, a company focusing on RNAi therapies, based in Pasadena, California, said: "The world saw this as a new panacea and the world goes crazy, the market goes crazy."

Then, about ten years ago, it dawned on people that getting the therapeutic strands of RNA into the right cells, so they could do their job of interfering with the target gene, was a greater challenge than first thought.

RNA is typically single stranded, so first researchers had to work out how to shield the therapeutic double-stranded RNA from degradation by the human immune system.

Second, they had to ensure that the RNA molecule would remain stable long enough to do its job.

After overcoming these hurdles, the researchers had to get the RNA into the target organ and cells, and once inside the cells, make sure the RNAi machinery was unleashed.

"The conundrum of RNA therapeutics has always been delivery," says Bo Rode Hansen, chief executive of Genevant Sciences, a company in Cambridge, Massachusetts, that focuses on delivery methods for RNAi.

It's too early to say that all these challenges have been overcome. But US and EU approval in 2018 of the first RNAi-based therapeutic product — patisiran, which treats a hereditary nerve condition caused by the accumulation of an aberrant protein — is an encouraging sign.

## HIDING IN PLAIN SIGHT

The human immune system is on high alert for double-stranded RNA, which is a hallmark of some pathogenic viruses.

"We know that double-stranded RNA is a really important danger signal," says Ken Pang, paediatrician and biomedical researcher at the Murdoch Children's Research Institute in Melbourne, Australia. "The innate immune system is finely attuned to seeing double-stranded RNA as a signal that viruses are present, and it starts mounting an immune response."

The first solution that the industry devised to tackle this was nanoscale spheres of lipid molecules to encapsulate the double-stranded RNA and hide it from the immune system. These lipid nanoparticles not only provide a shield, but also piggyback on existing cellular transport mechanisms to get the nanoparticle and its cargo into the cell. In the case of patisiran, the nanoparticle containing the double-stranded RNA is taken up by the receptor that normally transports low-density lipoproteins (LDLs) into the cell. Once inside the cell, the lipid nanoparticle dissolves, releasing its double-stranded RNA cargo into the cellular cytoplasm to silence the target gene.

A second approach is to engineer the double-stranded RNA so that it becomes invisible to the immune system. "The key concept here is that every base is chemically modified such that, while it's called an RNAi therapeutic, there's no pure RNA in the drug," says Kevin Fitzgerald, senior vice-president of research at Alnylam

Pharmaceuticals in Cambridge, Massachusetts, the company that makes patisiran. It turns out that if enough modifications are made to the RNA molecules, the immune system doesn't see them.

One advantage of this engineered double-stranded RNA is that it can be injected through the skin, whereas the lipid nanoparticle-based formulation has to be administered intravenously. What's more, the double-stranded formulation is stable at room temperature and has a long shelf life.

"You start to think what could you do with that in the less developed parts of the world, where practical cold chain is difficult, and dosing is difficult," says John Pasi, director of the Haemophilia Centre at Barts Health NHS Trust in London. "You start to think on a global scale about how this actually might have real value."

This engineered RNA approach is also being applied to a new RNAi therapeutic for haemophilia — a disease characterized by the lack of clotting agents in the blood. The drug, fitusiran, targets the production of antithrombin, a protein that inhibits clotting. By silencing a gene for antithrombin, fitusiran enables a rise in the production of the enzyme thrombin, which helps the blood to clot. Phase I and II clinical trials have shown that a fixed monthly dose of this subcutaneous treatment can significantly lower the number of some bleeding episodes that patients experience each year — down to zero in around two-thirds of patients. A phase II clinical trial found that 48% of people who were treated with a fixed monthly dose of this subcutaneous treatment experienced no bleeds at all during the observation period, according to Alnylam and pharmaceutical giant Sanofi.

Engineered RNA seems stable enough to survive even the hostile acidic environment of the stomach, possibly enabling oral delivery of RNAi therapeutics, say Alnylam.

Although the stability of double-stranded RNA means it is currently the dominant approach, single-stranded RNA is also a potential therapeutic candidate. Research suggests the other half of the double RNA strand isn't necessary for the interference action[1,2], and single-stranded RNA can also be chemically stabilized to survive inside the body.

Lipid nanoparticles and engineered RNA molecules aren't the only options being explored in the quest to overcome RNAi's delivery challenges. Anil Sood, co-director of the Center for RNA Interference and Non-Coding RNA at the MD Anderson Cancer Center in Houston, Texas, is leading a group that is investigating a polysaccharide-based nanoparticle called chitosan to deliver small lengths of RNA into the cell, with the aim of inhibiting the growth of prostate-cancer cells. The group is also looking at using a form of high-density lipoprotein (HDL) to create a nanoparticle-delivery platform.

"We feel like that that's a neat approach because it uses a system within the body that already exists," says Sood. Their research suggests reconstituted HDL could be an efficient way to get RNAi inside cells, and potentially better than the LDL-receptor uptake mechanism now used by patisiran's lipid nanoparticle. (High levels of LDL, sometimes called 'bad' cholesterol, can raise the risk of heart disease or stroke.)

HDL is taken into liver cells through a receptor called SR-B1. As well as being widespread in the liver, SR-B1 can also be found in tumour cells. That's significant because it has been difficult to find suitable RNAi targets for such cells.

## GOING BEYOND THE LIVER

If there was such a thing as low-hanging fruit in the fiendishly challenging landscape of RNAi, it would be diseases that involve the liver.

"The liver is a clearance organ, so between that and the kidneys it should be reasonably easy to get into those types of cells because the liver is trying to pull stuff out of the blood anyway," says Anzalone. Liver cells also have a well-known and characterized receptor — the LDL receptor — that can be used to transport sufficient quantities of RNA into those cells.

That has worked for lipid nanoparticles, but what about naked RNA, which would avoid the need for delivery mechanisms such as nanoparticles?

For this delivery challenge, researchers attach the engineered double-stranded RNA to a molecule called a ligand, which has a natural attraction to a particular type of receptor found on the surface of the target cell.

For the non-encapsulated version of patisiran, the RNA is joined to a ligand called *N*-acetylgalactosamine, or GalNAc, which has a high affinity for a liver hepatocyte receptor. The ligand drags the RNA to that receptor, and opens the door to the inside of the cell.

> ***"The conundrum of RNA therapeutics has always been delivery."***

With this development, delivery to liver-based targets is now straightforward. "It's an organ that we're all very confident that we can make safe and effective therapies in now," says Fitzgerald.

The eye is another organ for which delivery is relatively easy, although the method — direct injection into the fluid inside the eye — might make some a little squeamish. Fitzgerald says this approach is ideal because it delivers the treatment directly to the target cells of the eye. Alnylam's preclinical work on eye delivery suggests that doses as small as three micrograms can be effective and have a long-lasting effect.

Similarly, the central nervous system has a backdoor that enables researchers to bypass the blood–brain barrier — a tightly-packed conformation of endothelial cells that protects the brain from potentially harmful molecules circulating in the blood. Drugs can be injected directly into the intrathecal spaces in the spinal cord. The cerebrospinal fluid circulates the drug throughout the brain and spinal cord, and the ligand makes sure the drug reaches the target cells in the brain.

Targeting cancer with RNAi therapy is a daunting prospect, because the drug has to hit every single cell, says Fitzgerald. There's also the possibility that the tumour will adapt to the therapy by reducing the number of target receptors. This could be where the SR-B1 receptor — found in high numbers in many cancer cells — comes into its own.

## ESCAPING THE ENDOSOME

Once the double-stranded RNA has been encapsulated or stabilized, found its way to its cellular target and been transported inside the cell, there's still one last hurdle to overcome. When the cell takes in the double-stranded RNA complex, it forms a membrane around the RNA that then buds off — a process called endocytosis.

The double-stranded RNA must escape the membrane so that it can then reach the cellular machinery and interact with and silence its target gene.

Pang and his colleagues have identified a receptor, called SIDT2 in mammals, which is responsible for the release of viral double-stranded RNA into the cytoplasm as part of the normal immune response to viral infection[3]. But for some reason, the receptor doesn't seem to free RNAi therapeutics from their membrane vessels, or endosomes, in the same way.

Anzalone says that Arrowhead Pharmaceutical's approach relies on the fact that the engineered RNA is stable enough to survive inside the endosomes until a small number leak out naturally. RNAi is so efficient that only a few molecules need to leak out into the cellular fluid to produce the desired effect, he says.

Fitzgerald says endosomal escape can actually work in RNAi's favour, because it allows the double-stranded RNA to be released slowly, and the effect, therefore, to be extended and dosing to be less frequent.

## DELIVERING ON THE POSSIBILITIES

Delivery has been such a significant challenge that those working in the area of RNAi therapeutics might be forgiven for admitting defeat. And there's still a long, hard road ahead, says Fitzgerald. "Each organ is a little bit different and the challenges of delivery are a little bit different," he says. "As we go into each new organ, there will be some learning to do."

Despite this, enough researchers have clung to the vision of RNAi as a potent new therapeutic modality, and refused to give up. Their persistence and years of toil are paying off, says Anzalone, delivering treatments that "can change millions of lives". ■

---

**Bianca Nogrady** *is a freelance science writer in Sydney, Australia.*

1. Lima, W. F. *et al. Cell* **150**, 883–894 (2012).
2. Yu, D. *et al. Cell* **150**, 895–908 (2012).
3. Nguyen, T. A. *et al. Immunity* **47**, 498–509.e6 (2017).

# INJECTION OF HOPE

*Messenger RNA could help to boost immunity against cancer, influenza and much more.*

**Brad Kremer received an experimental vaccine for his cancer.**

KATYE MARTENS BRIER

**BY ELIE DOLGIN**

Brad Kremer had waited months to receive an experimental cancer vaccine called BNT122, during which time the melanoma on his skin had spread to his liver and spine. His back pain was getting worse, he was rapidly losing weight and new cancerous lesions kept appearing on his left thigh. "It was very scary," says Kremer, a 52-year-old sales representative from Acton, Massachusetts.

But within weeks of his first injection in March, Kremer could see that the vaccine was working. The coin-sized melanoma spots that popped up from his skin were now flat discolourations measuring millimetres across. "I was actually witnessing the cancer cells shrinking before my eyes," he says. Several doses later, his appetite has returned, his back pain has subsided and scans show that his cancer is continuing to retreat.

Kremer's dramatic response exemplifies the medical potential of vaccines built on messenger RNA. In this method, strings of lab-synthesized nucleotides train the immune system to recognize and destroy disease-causing agents — be they cancer cells or infectious viruses.

Other ways of making vaccines can achieve the same therapeutic objective. But the potency, versatility, speed of manufacturing and low cost of mRNA make it an attractive platform for the rapid development and large-scale production of new or custom-made vaccines.

Early clinical results have demonstrated the technology's promise. Researchers at BioNTech in Mainz, Germany, the manufacturer of the cancer vaccine that Kremer is receiving, reported in 2017 that all of the first 13 people with advanced-stage melanoma to receive the personalized immunotherapy — which is tailor-made to match the genetic profile of each person's cancer — showed elevated immunity against the mutated bits of their tumours. As a result, these patients' risk of developing new metastatic lesions was significantly reduced[1]. For viral diseases, prophylactic vaccine candidates against rabies[2] and pandemic influenza[3] have each proved safe and induced protective antibody responses in healthy volunteers. In both cases, however, the antiviral effects waned after less than a year, suggesting that improvements are needed to provide more robust and long-lasting immunity.

"There's a lot of potential here," says John Mascola, director of the Vaccine Research Center at the US National Institutes of Health in Bethesda, Maryland. "It's still early in the development of these vaccines, but the platform has shown proof of concept."

## MADE TO MEASURE

Moderna Therapeutics is racing to develop mRNA vaccines further. In 2018, company executives cut the ribbon on a US$130-million, two-storey, football-pitch-sized manufacturing plant in Norwood, Massachusetts, about one hour's drive from the company's headquarters in Cambridge. Technicians at the new site synthesize and formulate mRNA for all of Moderna's early-stage clinical trials, including for an ongoing investigation of mRNA-4157, a personalized cancer vaccine that, like BioNTech's BNT122, has shown preliminary signs of anti-tumour activity in people with cancer.

Along with other companies testing the same strategy, Moderna starts the process of making its personalized treatment by taking a pair of genetic profiles from each individual: one from a biopsy of the tumour, the other from a vial of healthy blood cells. Algorithms compare the DNA sequences of the two samples and produce a list of 34 targets, each encoding a different mutant protein expressed by the cancer that is predicted to be useful in training the immune system to attack the disease.

The technicians then take those digital sequences from the computer model and turn them into physical products. First, they codify the cancer-specific mutations into a ring-shaped DNA molecule called a plasmid. They then convert the DNA into strands of mRNA before the therapeutic nucleotides are finally coated in lipid nanoparticles to make them more stable in the body.

Throughout this month-long manufacturing process, each patient's tailor-made vaccine is assigned a colour — red, blue, yellow or green — and technicians wear colour-coded hairnets, work behind colour-matched lines on the laboratory floor and use only colour-coordinated equipment, all as a visual precaution against mix-ups. "We need to make sure it's one batch per patient all the way," explains Jennifer White, head of quality assurance at the site.

At the American Society of Clinical Oncology's annual meeting in June, Moderna reported first-in-human clinical data showing that mRNA-4157 can generate mutation-specific immune responses in people with cancer. And when administered with a checkpoint inhibitor — an antibody drug designed to further ramp up the body's cancer-fighting immune activity — the therapy also shrank tumours in 6 of 20 participants with metastatic disease. Moderna and its development partner, Merck, based in Kenilworth, New Jersey, launched a randomized, 150-person follow-up study in July.

As personalized mRNA vaccines go through trials, Moderna, BioNTech and CureVac, based in Tübingen, Germany, are simultaneously developing off-the-shelf vaccine candidates as well. These ready-made vaccines are not as immunogenic as the most potent customized vaccines for people with highly mutated cancers, but they are potentially suitable for everyone. There is no lengthy customization process; no long waits, often of a month or more; and no added labour and manufacturing costs. They are ready for anyone who needs them.

## A NEED FOR SPEED

Oncology is one important area for mRNA vaccine manufacturers, but they are also developing products to tackle infectious diseases. They are taking advantage of the platform's ease of manufacture to create rapid-turnaround products. And they are deploying the technology to combat a handful of viruses that have remained impervious to conventional vaccine strategies.

One such target is cytomegalovirus (CMV), the most common infectious cause of neurological defects in newborns in the developed world. A vaccine is desperately needed to prevent pregnant women from passing the virus to their developing fetuses. But vaccine makers have struggled to recreate the virus' pentameric complex — a bundle of five proteins that mediate entry and exit into human cells — in a way that generates a robust immune response when introduced in the body.

"You can't make five different things in a vat, purify them and then try to put them together in the lab," says Tal Zaks, Moderna's chief medical officer. "They have to be put together within the cell," he says, "and mRNA allows researchers to do that."

Moderna scientists demonstrated this last year, creating a multi-sequence mRNA vaccine that prompted cells transfected with the mRNA to express the full pentamer on their surfaces, eliciting protective antibody responses in immunized mice and monkeys[4]. And last month, the company disclosed that participants in a phase I human trial experienced dose-dependent increases in antibody levels as well. Other prophylactic vaccines in Moderna's pipeline with promising early clinical data include one for respiratory syncytial virus, a common cause of airway inflammation in infants. And in the case of the company's mRNA-1653, a dual vaccine against two other recalcitrant respiratory viruses, metapneumovirus and a type of parainfluenza, which are also responsible for severe lung infections, the success of Moderna's early trials shows that "you can actually now do combinations", Zaks says.

Moderna is also working on vaccines to tackle emerging infectious diseases such as avian influenza and Zika virus, for which the speed of mRNA manufacturing could be beneficial in the event of a pandemic. "One of the greatest advantages of this mRNA strategy is just how fast you can go from a nucleotide sequence to a vaccine product," says Justin Richner, a vaccine researcher at the University of Illinois College of Medicine in Chicago, who has collaborated with Moderna in the past.

The current speed record was set in 2013 in response to an influenza outbreak in China: it took scientists at Novartis just eight days to make a vaccine candidate. Health officials at the Chinese Center for Disease Control and Prevention in Beijing posted gene sequences from the virus on a data-sharing platform on a Sunday in late March[5]. By the following Sunday, Andy Geall and his team at Novartis's vaccine unit in Cambridge, Massachusetts, were already running validation experiments of their mRNA vaccine in hamster kidney cells.

"It happened in real time the moment that sequence was available," says Geall, who is now head of formulations, analytics and chemistry at Avidity Biosciences in La Jolla, California. For a conventional vaccine, the same process could take six months or more, he adds.

> "YOU CAN USE MESSENGER RNA VACCINES FOR PRETTY MUCH EVERYTHING."

Geall continues to consult with companies and academics working on mRNA vaccines. He has recently started to go beyond viruses to other pathogens. Last year, for example, he joined forces with clinical immunologist Richard Bucala's team at Yale School of Medicine in New Haven, Connecticut, and successfully used mRNA to vaccinate mice against malaria[6], which is caused by a single-celled parasite.

"The whole platform is very, very flexible," says Norbert Pardi, who studies infectious diseases at the University of Pennsylvania's Perelman School of Medicine in Philadelphia. He is currently working on vaccines for malaria as well as HIV, hepatitis C and several other viral diseases. "You can use mRNA vaccines for pretty much everything," he says.

## TAKE ANOTHER SHOT

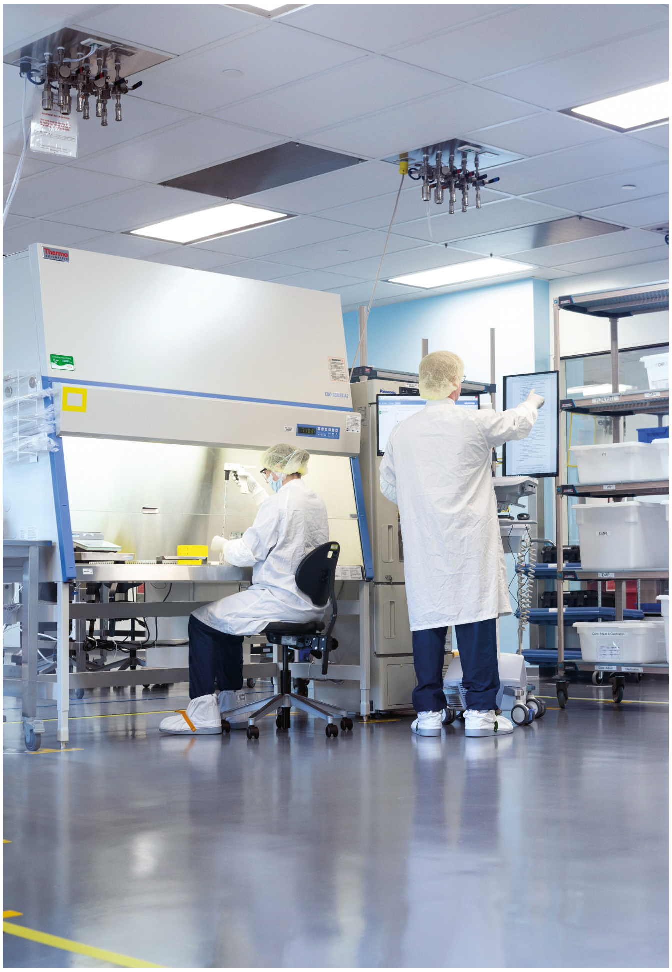
Developing mRNA vaccines is not always straightforward, however. Moderna's initial candidate vaccine for Zika virus, for example, was well tolerated in people but failed to provoke much of an immune response. With funding from the US government, Moderna went back to the lab, optimized the vaccine sequence and developed another candidate that is, according to Zaks, "at least 20 times more potent" than the first-generation product in mice and monkey testing.

The first clinical-stage mRNA vaccine from CureVac was also disappointing. It was a rabies vaccine[2] that could induce antibody responses in some study participants, but only when administered through needle-free devices — and even then, only around two-thirds of vaccinated people achieved the recommended level of antiviral therapy.

CureVac has since altered its delivery platform and restarted human trials with a vaccine candidate that is encased in lipid nanoparticles. As the company's scientists reported in 2017, this change enhances the cellular uptake of the mRNA sequences, so the same level of antibody and T-cell responses can be achieved in mice and monkeys using a tiny fraction of the dose[7].

But starting clinical development again cost the company valuable time in the race to bring its products to market. For chief technology officer Mariola Fotin-Mleczek, this was an important lesson learnt the hard way. "Formulation makes a big impact," she says.

What goes into the mRNA sequence matters a great deal, too. Some companies and non-profit groups are pushing ahead with


MODERNA

Technicians at Moderna make a personalized cancer vaccine.

'self-amplifying' mRNA vaccines. These include everything found in a conventional mRNA vaccine, plus all the genes that encode the RNA replication machinery. This combination of genes allows one strand of synthetic mRNA to generate thousands of copies of the mRNA of interest, says Jeffrey Ulmer, head of preclinical research for vaccines at GlaxoSmithKline in Rockville, Maryland. This amplification effect, he points out, "gives you the potential to have a much more potent immune response with the same dose of RNA — or alternatively, you need much less RNA to get an equivalent level of immune protection".

For instance, BioNTech scientists, working with immunologists from Imperial College London, reported that mice could be fully protected from influenza using just 1.25 micrograms of self-amplifying mRNA — a small fraction of the 80 micrograms of a conventional mRNA vaccine needed to produce the same effect[8]. Late last year, scientists at Imperial collaborated with the Coalition for Epidemic Preparedness Innovations, a global public–private partnership based in Oslo, to advance its self-amplifying mRNA platform to produce vaccines against rabies, the Marburg virus and H1N1 influenza.

Beyond vaccines, mRNA technologies can be used to produce the missing or defective proteins responsible for all manner of diseases. Many companies, including Translate Bio of Lexington, Massachusetts, are pushing ahead with this kind of restorative therapeutic strategy. But as Translate Bio's chief scientific officer, Richard Wooster, points out, targeting corrective mRNA drugs to the desired organs and cell types in the body muscle, say, or to cancer cells, is extremely difficult. In addition, scientists are still struggling to devise chemical formulations of mRNA that avoid unwanted immune reactions to the molecules.

Vaccines are not hampered by these technical complexities. "With vaccination, you're getting a systemic effect, so you don't necessarily need to have specific tissue targeting," says Wooster, whose company signed a deal in 2018 with Sanofi Pasteur of Paris to develop mRNA vaccines for up to five pathogens. Furthermore, because the goal with any vaccine is to elicit immune activity, the tendency for lab-made mRNA to trigger the immune system — and thus serve as its own immunological enhancer — becomes a feature, not a problem.

Most of the mRNA vaccines developed so far have focused on cancer and infectious diseases. But BioNTech also holds the patent rights to an mRNA-vaccination platform designed to protect against allergens such as grass pollen and house dust mites. "I think it would be a very elegant way to prevent allergies," says Richard Weiss, an immunologist at the University of Salzburg, Austria, who helped to develop the technology. His team showed that mRNA-based immunization can completely protect mice from developing allergies against timothy grass, a common cause of hay fever[9].

> "I WAS ACTUALLY WITNESSING THE CANCER CELLS SHRINKING BEFORE MY EYES."

All companies have to set strategic priorities, however, and for now BioNTech's remain squarely rooted in oncology. Last month, its scientists reported at a cancer immunotherapy conference in Paris that the company's most advanced off-the-shelf vaccine candidate helped to shrink or stabilize melanoma in 19 of 42 early-trial participants. In continued follow-up investigations from the firm's first-in-humans study of its personalized cancer vaccine, all the people who had previously responded favourably were still relapse-free up to 41 months after treatment. And those responses came without the addition of a checkpoint inhibitor.

BioNTech is banking on the idea that the combination that Kremer is receiving for his metastasized melanoma will yield even better results, especially now that most of the kinks have been ironed out. Back in February, Kremer was supposed to receive his first dose of personalized vaccine shortly before taking a holiday to the Florida Keys with his wife and two teenage daughters. But the treatment, which BioNTech manufactures in Germany and ships to patients around the world, was held up at customs. The dry ice used to chill the vaccine vaporized, so the product thawed out and had to be discarded.

Kremer says he maintained a positive outlook as he waited three weeks for the next shipment of vaccine to arrive. His wife, Ginny, was less patient. "She was very upset," Kremer says. "I don't think she slept for two months."

BioNTech has since developed systems to ensure that its personalized vaccines reach people as quickly as possible — a crucial feature for any cancer treatment when time is of the essence. Fortunately for Kremer, the delay didn't impact his prognosis. "He's had a remarkable response," says cancer immunologist Ryan Sullivan, who is treating Kremer at Massachusetts General Hospital in Boston. And although "there are shipping concerns that can happen and not everyone responds", Sullivan adds, "with cases like Brad's, it's easy to get excited about this type of strategy." ■

**Elie Dolgin** *is a science journalist in Somerville, Massachusetts.*

1. Sahin, U. *et al. Nature* **547**, 222–226 (2017).
2. Alberer, M. *et al. Lancet* **390**, 1511–1520 (2017).
3. Feldman, R. A. *et al. Vaccine* **37**, 3326–3334 (2019).
4. John, S. *et al. Vaccine* **36**, 1689–1699 (2018).
5. Hekele, A. *et al. Emerg. Microb. Infect.* **2**, e52 (2013).
6. Baeza Garcia, A. *et al. Nature Commun.* **9**, 2714 (2018).
7. Lutz, J. *et al. npj Vaccines* **2**, 29 (2017).
8. Vogel, A. B. *et al. Mol. Ther.* **26**, 446–455 (2018).
9. Hattinger, E. *et al. J. Immunol. Res.* **2015**, 797421 (2015).


NEIL WEBB

HEART DISEASE

# RNA therapy sets sail for the heart

*Researchers hope that understanding the many roles of non-coding RNA in heart health and cardiovascular disease could deliver a therapeutic breakthrough.*

**BY CHRIS WOOLSTON**

When Stefanie Dimmeler isolates strands of RNA in her laboratory, she's not just handling an interesting piece of genetics. She's searching for a possible advance to fight the world's number one cause of death.

In the past decade, scientists such as Dimmeler have shown that certain types of RNA help to regulate the basic mechanisms of cardiovascular disease, including the formation of cholesterol, the build-up of plaques and the death of cells after a heart attack.

By learning more about the roles of RNA in cardiovascular disease, researchers hope to harness the molecules to prevent and even reverse key steps in the process.

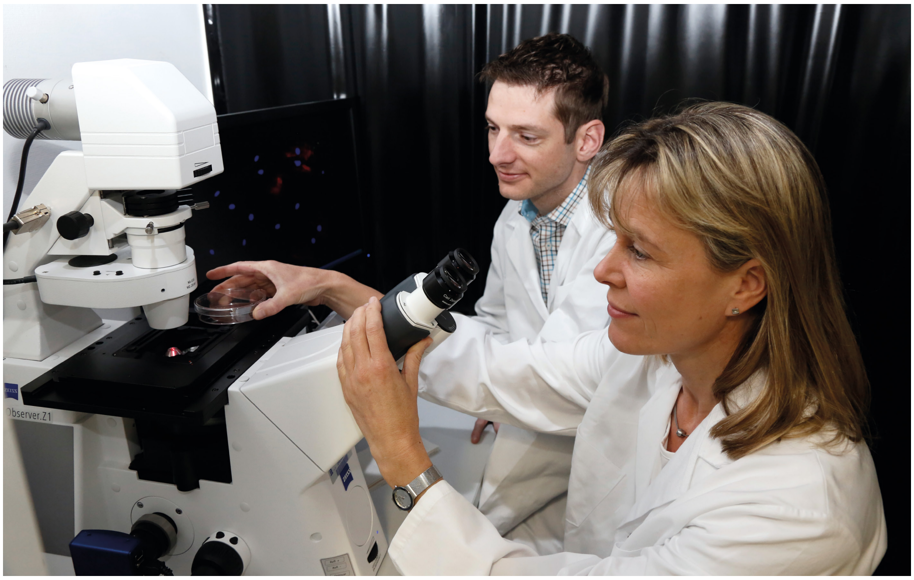
Dimmeler, a specialist in cardiovascular regeneration at Goethe University Frankfurt in Germany, describes RNA as a "biologically and mechanistically attractive target".

## BENEFITS AND RISKS

Because cardiovascular disease is often caused by a complex process that occurs slowly over decades, any therapy targeting RNA would have to deliver long-lasting results with minimal side effects, a difficult balance to strike. The near ubiquity of heart disease also raises the stakes. Researchers have to think carefully about who — if anyone — could benefit from this potentially risky treatment. "If we can get something going with RNA, the impact could be major," she says.

Most of the focus is on a class of RNA that doesn't directly code for proteins. Once thought of as meaningless genetic noise, such non-coding RNA is now known to regulate the expression of proteins throughout the body, including many of the key proteins involved in heart disease. In particular, non-coding microRNAs — strands of 21–23 nucleotides — are thought to affect the protein output of more than 60% of human genes, says Carlos Fernandez-Hernando, a pathologist at Yale University in New Haven, Connecticut.

Some of the non-coding RNA molecules that have already been isolated, identified and tested in the laboratory have created



Stefanie Dimmeler at Goethe University Frankfurt, Germany, investigates RNA that affects heart health.

COURTESY OF STEFANIE DIMMELER

real cause for optimism, Fernandez-Hernando says. Long non-coding RNA, which is made up of more than 200 nucleotides and can both promote and suppress the expression of proteins, intrigues researchers with its diversity and flexibility. But when it comes to therapeutic potential in the short term, scientists such as Fernandez-Hernando have focused mainly on microRNAs. These strands have just one main function: gumming up the production of proteins by binding to the transcription machinery in a cell.

Fernandez-Hernando's lab is especially interested in suppressing miR-33, a type of microRNA that has an important role in cholesterol metabolism. Among other functions, the molecule slows the production of high density lipoprotein (HDL) cholesterol, known as 'good' cholesterol because it helps transport low density lipoprotein (LDL) cholesterol to the liver, where it can be broken down and eliminated. Fernandez-Hernando and his colleagues have shown[1] that levels of HDL cholesterol in mice surge when the animals are treated with DNA sequences called antisense oligonucleotides that target and destroy miR-33.

> *"MicroRNA is wonderful because you ought to be able to target an entire pathway."*

Fernandez-Hernando explains that miR-33 inhibits at least ten genes involved in producing HDL cholesterol. The molecule's effect on any particular gene is fairly mild, he says, reducing the expression of a protein by just 5–10%. But because the microRNA affects different parts of the same biological process, suppressing the molecule can have a substantial impact on the final outcome. "MicroRNA is wonderful because you ought to be able to target an entire pathway," he says. "The effect on the output could be lot stronger than if you targeted a single gene. There's a lot of excitement for a potentially powerful technique."

## CARDIAC AGEING

That same excitement has spurred Dimmeler and her team in Frankfurt to investigate a suite of microRNAs that threaten heart health at the most fundamental levels. The microRNA miR-34, for example, promotes cardiac ageing and cell death after a heart attack. Just as Fernandez-Hernando hopes to boost good cholesterol by suppressing miR-33, Dimmeler thinks that lowering the production of miR-34 could promote healing after a heart attack.

Another molecule in Dimmeler's sights is miR-92A, a microRNA that induces cell death in cells deprived of oxygen after a heart attack. Dimmeler and her team showed that inhibiting miR-92A significantly improved heart function in mice and pigs with sudden insufficient blood supply to the heart, or acute ischaemia[2]. "There's also unpublished data showing that miR-92A can improve heart function in diabetic pigs with chronic ischaemia," she says. "It could be a very interesting target for heart failure."

Other microRNAs work the other way, by encouraging regeneration of heart tissue. Instead of suppressing these targets, researchers are hoping to make them more abundant. In 2012, scientists showed that over expression of miR-199 enhanced the proliferation of heart cells in mice[3]. In practice, such an approach could trigger an unhealthy overload of heart cells. Dimmeler says that any treatment that uses miR-199 to encourage regeneration of damaged tissue would have to be carefully titrated and given for a short time.

Other obstacles might prove tougher to overcome. Because microRNAs can attach to many different genes, unintended consequences are often a serious concern, says Andrew Baker, a molecular biologist at the University of Edinburgh, UK. Studies of miR-21, a microRNA associated with plaque build-up in arteries, have found it is seven times more abundant in peripheral arteries clogged with atherosclerosis than in healthy arteries[4]. But broad, unfocused attempts to suppress the molecule will almost certainly hit other targets with unpredictable results. Baker says "miR-21 is expressed in virtually every cell". "You need to target the therapy to a particular cell type," he adds.

Scientists are exploring several strategies for delivering RNA therapies to the intended cardiovascular targets, including piggybacking microRNA onto an adenovirus vector — commonly used to insert genes into cells — and using catheters to deliver the package to specific tissues. Another option is using microRNA inhibitors that could be activated by precisely aimed beams of light. Dimmeler notes that using light can be effective for surface tissues but is much more challenging for internal organs such as the heart. Baker and his colleagues are exploring ways to deliver miRNA-based therapies directly to veins exposed during surgical grafts. "Having access to the tissue is a real bonus," Baker says. "It also presents some challenges because you don't have access to the tissue for very long."

If and when such technical hurdles are overcome, the central question remains: who might benefit from cardiovascular therapy that targets RNA? Baker suspects that it could be most helpful for people at very high risk of heart disease, such as those with unstable plaques or early-stage heart failure. Targeting a process such as the slow build-up of plaque in arteries, a nearly universal occurrence that develops throughout a lifetime, would require further breakthroughs and a deeper understanding of the risks of manipulating RNA for decades at a time. "There are a lot of unknowns about such long-term treatment," Baker says. "That's not to say that it can't be done."

The uncertainties should help to spur more basic research in the future. In addition to fine-tuning approaches using known microRNAs, Baker says that researchers should continue looking for new candidates. It's quite possible, he says, that the best targets for cardiovascular disease have yet to be discovered. "There are a lot of non-coding RNAs in the human genome that nobody knows anything about," he says. Until all of RNA strands at the core of heart disease have been identified, the full potential of the RNA approach will remain unknown.

It might turn out that RNA-based therapies never add much firepower to the fight against cardiovascular disease, Baker says. But with so many potential targets and so much on the line, it's clearly worth a shot. ■

---

**Chris Woolston** *is a freelance writer in Billings, Montana.*

1. Rotllan, N. *et al. Arter. Thromb. Vasc. Biol.* **33**, 1973–1977 (2013).
2. Hinkel, R. *et al. Circulation* **128**, 1066–1075 (2013).
3. Eulalio, A. *et al. Nature* **492**, 376–381 (2012).
4. Hung, J. *et al. Front. Physiol.* https://doi.org/10.3389/fphys.2018.01655 (2018).

## PERSPECTIVE

# Eclipsed by CRISPR

As a source of potential treatments, gene editing gets all the attention. It's time for scientists to shout about RNA therapies, says **Lorna Harries**.

The merits and risks of gene editing and gene-manipulation technologies such as CRISPR–Cas9 have been hot topics for debate ever since such techniques were developed. Ask anyone on the street about gene editing and the chances are that most will have heard of it. There might be controversy over the benefits of using gene editing to tackle conditions and the ethical challenges it poses, but the public will probably be aware of the field's existence and potential. Ask the same people about RNA therapies (or even just RNA) and you will probably be greeted with a blank stare.

Why is this the case? Several RNA therapies have already been licensed for use in the clinic, yet the development of CRISPR–Cas9 in a clinical context is still in its infancy. The discovery of RNA was concomitant with that of DNA, but it has yet to appear on the public radar, despite the opportunities that it offers as a therapeutic tool.

One reason that RNA-based approaches are the lesser-known and seldom-seen sibling of DNA therapies might be the evident complexity of RNA biology, and the perceived difficulties in talking about it to a non-scientific audience.

> WITH SAFE AND EFFECTIVE TREATMENTS, GENETIC DISEASES THAT ARE INCURABLE NEED NOT BE A LIFE SENTENCE.

As a scientist and educator, I am consistently met with the perception that RNA biology is too difficult to understand, and is therefore best left alone. It is true that it is a complex area, but so are genetics and epigenetics. Yet both are represented in and discussed by the media in a way that RNA biology is not. It is down to scientists and communicators to make the subject accessible and easy to understand. It's not impossible — we just need to engage with the task and to give it priority. I find that the public is actually very receptive to information on the topic, and that with understanding often comes a willingness to embrace things.

The idea that you can change a DNA sequence by using a one-stop therapy to repair a genetic mutation is extremely attractive. But such interventions are difficult to revert, and have the potential to be transmitted to future generations. This area is one in which RNA therapies might have the upper hand, representing a more digestible solution to correcting genetic conditions for clinicians and the public. Unlike gene editing, RNA therapies do not alter the actual sequence of a mutated gene, but instead alter its output. Those changes are temporary, not permanent. As much as 50% of diseases in people might arise from changes in RNA transcription, processing or turnover. Increasingly, we are realizing that the risk of developing many common long-term conditions is also influenced by changes in the expression of genes, rather than in the structures of the proteins that they encode.

RNA therapies might be relatively new, but already several such drugs have made it to the clinic. These include: patisiran for hereditary transthyretin amyloidosis, a build-up of the protein amyloid in nerves and organs; eteplirsen for certain types of Duchenne muscular dystrophy, an inherited form of progressive muscle degeneration; and nusinersen for spinal muscular atrophy, a rare genetic neuromuscular condition. Before the advent of these drugs, all three conditions were life threatening, or potentially so.

Such drugs work by changing how RNA-encoded messages sent from mutated genes are produced, processed or degraded. This can restore the number of messages that are produced, or alter the nature of those messages so that they more closely resemble those from genes without mutations.

In many ways, RNA therapies are not so different from other existing medications. They are prescribed and taken regularly (but perhaps less often than are conventional drugs), and their effects do not last indefinitely.

As well as the challenges imposed by RNA therapies in terms of their delivery, targeting and efficacy, there are some less tangible barriers to their use.

First, there might be psychological barriers to their uptake. Any new treatment can seem frightening to patients and their families. Even medical professionals sometimes prefer to stick with less effective, but more familiar, approaches to managing conditions. There is work to be done in making RNA therapies less alien and more akin to conventional medications, but this is best done by communicating the vast health benefits that they might bring. The first applications of such drugs have been in rare genetic diseases. Parents of children with these conditions have often run out of conventional treatment options. They might also be better connected to medical professionals and therefore be more willing to try fresh approaches.

Second, RNA therapies are more expensive than many drugs that are already on the market. The UK National Institute for Health and Care Excellence, which approves treatments for use by the UK National Health Service on the basis of cost-effectiveness, currently quotes a cost of £450,000 (US$553,000) for the first year of treating spinal muscular atrophy with nusinersen, and then £225,000 for each subsequent year. Advances on the scale of RNA therapies do not come cheap, and we should not underestimate the value of an effective therapy to the family of a child with spinal muscular atrophy.

Finally, by and large, these medications are best for tackling conditions with a common genetic aetiology. In an era of personalized medicine, it might not be cost-effective to design a specific intervention for individual mutations. However, many inherited diseases do share specific types of mutation, and these are ripe for the development of RNA therapies.

This fresh approach to tackling genetic diseases has the potential to create a sea change in how we think about treating people with inherited conditions. With safe and effective treatments that really work, genetic diseases that are incurable at present need not be a life sentence in the future. The barriers that remain are not insurmountable. We still have work to do, but the future is bright for RNA therapies. ■

---

**Lorna Harries** *is professor of molecular genetics at the University of Exeter, UK.*
*e-mail: l.w.harries@exeter.ac.uk*

THERAPEUTICS

# A tale of two drugs

*A rare genetic disease became a focus for two pharmaceutical companies that specialize in RNA therapies. Now, treatments for other hereditary conditions are within the sights of the industry.*

MARCUS HESSENBERG/BARCROFT IMAGES/BARCROFT MEDIA VIA GETTY

**BY LIAM DREW**

In 2018, US and European regulators approved two RNA-targeted drugs for the rare genetic disease hereditary transthyretin (hATTR) amyloidosis, offering hope to around 50,000 people with the condition worldwide.

The disease is caused by mutations in the gene that encodes transthyretin, a protein in blood that is made by the liver. Such mutations lead to a decrease in the solubility of transthyretin, so that, instead of being transported through the body, the protein forms clumps around nerve fibres and in the cardiovascular system, progressively eroding tissue function.

The drugs that were approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) enter liver cells to thwart transthyretin production. Each binds to the messenger RNA that encodes the mutated protein. But these drugs are otherwise distinct, and represent the two most advanced strategies for targeting RNA therapeutically.

One of the drugs, inotersen (Tegsedi), is an antisense oligonucleotide (ASO). Typically, ASOs are short single strands of modified DNA that are designed to bind to a specific mRNA. They thereby suppress that mRNA's ability to guide protein synthesis, and often lead to destruction of the mRNA.

ASOs were discovered more than 40 years ago, and the developers of inotersen, Ionis Pharmaceuticals, founded in 1989 and based in Carlsbad, California, led the long process of chemically modifying DNA to convert ASOs into functional drugs. Inotersen was only the fifth medication in this class to gain clinical approval.

The other approved treatment for hATTR is patisiran (Onpattro), which was developed by Alnylam Pharmaceuticals in Cambridge, Massachusetts. Patisiran comprises a short stretch of double-stranded RNA known as a short interfering RNA (siRNA) and is the fruit of Nobel-prizewinning research that showed plants and animals use a mechanism called RNA interference (RNAi) to control the abundance of mRNA. In RNAi, cells produce siRNAs that are complementary to sections of specific longer mRNAs. When an siRNA binds to its target mRNA, it catalyses the destruction of that molecule.

Alnylam was founded in 2002 with the goal of hijacking this pathway to make RNA therapies. Patisiran is the first RNAi-based drug to be approved.

When recounting the histories of their respective companies, Stanley Crooke, founder and chief executive of Ionis, and Kevin Fitzgerald, senior vice-president of research at Alnylam, tell similar stories. Biological technologies, they say, go through three phases during development. First is what Fitzgerald calls the "hope and promise" phase, in which investors bet big on the technology. But this is almost always followed by a phase of slow progress and growing scepticism, during which funding is harder to come by and developers must innovate to deliver on initial promises.

The third phase — if reached — begins with the first promising data in human trials and proceeds to drugs getting approved for clinical use. According to Crooke, when this happens, people start to argue over who came up with the technology in the first place.

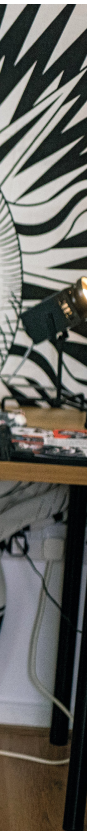
RNA therapies seem to be securely in the third phase. What remains to be seen is how widely they will affect clinical practice.

## FIRST, THE LIVER

To work, RNA therapies must meet a number of criteria. Most importantly, ASOs and RNAi drugs should be stable enough to persist in the body for a useful length of time — around a day — to allow for a reasonable dosing regimen. On administration, they must reach their target organs, enter the target cells and then bind tightly to target mRNAs. The ASOs that existed when Ionis was founded in 1989 had none of these properties. And the first human trials in the 1990s revealed that they had serious side effects. In 2002, the development of RNAi drugs was facing many of the same issues.

In 1998, Ionis got clinical approval for fomivirsen (Vitravene) — an ASO based on early chemistry that treated CMV retinitis in people with AIDS. Even so, throughout the 1990s the company focused mainly on modifying DNA for use in ASOs.

By tweaking the chemical structure of DNA, researchers were able to make ASOs from synthetic analogues of DNA that retained the molecule's crucial property of binding to specific RNA targets, on the basis of Watson–Crick base-pairing, but that otherwise behaved differently from natural DNA. Gradually,



**Ellie Darby-Prangnell has a genetic condition called spinal muscular atrophy.**

molecules emerged that were more resistant to nucleases, enzymes that form part of the body's mechanism for degrading stray nucleic acids. This drastically increased the biological half-life of ASOs. Other improvements included binding to target RNAs more tightly and generating fewer inflammatory side effects.

David Corey, a pharmacologist at the University of Texas Southwestern Medical Center in Dallas, says that scepticism about ASOs grew during the 1990s, but at the start of the 2000s, the science suddenly became convincing. Corey attributes this to two major factors: the development of innovative classes of ASO, and the realization that researchers needed to focus on where ASOs accumulate in the body — the liver.

Alnylam discovered that the same was true of siRNA therapies, and therefore focused its attention on developing two strategies for getting these drugs into liver cells.

The first approach involves packaging siRNAs into vesicles comprising lipid nanoparticles, which mimic naturally occurring lipid droplets that are taken up by liver cells through the low-density lipoprotein (LDL) receptor. The second requires conjugating siRNA to a small molecule called *N*-acetylgalactosamine (GalNAc) that binds to and enters liver cells through a receptor that is selectively and highly expressed on such cells, a tactic that Ionis also used to further hone ASO delivery to the liver.

RNA therapies do not yet work as orally administered pills. They must be given by subcutaneous injection or, for lipid-nanoparticle formulations, by intravenous infusion — a process that requires a health-care professional. But the drugs now act for weeks or months, potentially compensating for these more cumbersome administration routes. Focus on the liver yielded the second ASO to gain FDA approval — 14 years after the first. Mipomersen (Kynamro), developed by Ionis, treats homozygous familial hypercholesterolaemia, an inherited disease that elevates cholesterol levels in the blood by reducing the production of a protein called apolipoprotein B in the liver. This was a very important drug, says Corey, because its intended effects on apolipoprotein B and cholesterol were directly observed in humans.

Although the main goal was to treat the genetic condition, Ionis saw mipomersen as a drug that might ultimately be used to tackle cholesterol levels in the general population. It touted the ASO as a future competitor of the cholesterol-lowering statin atorvastatin (Lipitor): infrequent, long-lasting injections that would replace daily pills.

But mipomersen can lead to liver toxicity and serious cardiovascular side effects. In 2012, it was refused a licence by the EMA, and clinicians in the United States favoured lomitapide (Juxtapid), a small-molecule drug developed by Aegerion Pharmaceuticals in Cambridge, Massachusetts, that lowers apolipoprotein B levels by a different mechanism.

Side effects and ease of administration are important factors in drug development. People with conditions such as hATTR amyloidosis and other rare genetic disorders, for whom there might be no other treatment options, are often more willing to tolerate drugs with side effects. They also visit doctors regularly, which could explain why they are often amenable to more-involved drug-delivery routes.

Inotersen and patisiran are far from being perfect drugs. For example, inotersen is more conveniently given by subcutaneous injection, but recipients must be monitored for a decrease in blood platelet counts — a serious potential side effect. Patisiran has a better side-effect profile, but requires intravenous administration in a clinic, and those given the drug must take steroids to suppress infusion-related reactions. The relative success of these drugs will not depend simply on their efficacy, but also on patients' experiences of taking them.

## MASS APPEAL

Using RNA therapies to treat genetic liver conditions seems to be an approach that works. In May, Ionis gained approval from the EMA (but not the FDA, when it considered the application in August 2018) for its third liver-acting ASO, volanesorsen (Waylivra), to treat a hereditary disorder that elevates levels of fats called triglycerides in the blood. And in August, the FDA granted givosiran priority-review status. The drug is an siRNA developed by Alnylam for acute hepatic porphyria, another inherited liver disease.

But researchers in the field hope that liver-targeting RNA therapies will soon be used to treat more-common disorders. There are more than ten ongoing phase III clinical trials of such therapies for liver-related diseases. Ionis, for example, has developed candidate ASO-based treatments for thrombosis, which target a blood-clotting factor made in the liver, as well as an ASO that aims to decrease levels of apolipoprotein A — high levels of which are a risk factor for cardiovascular disease, and for which no lowering strategy is available.

It is probable, however, that a GalNAc-conjugated siRNA called inclisiran will be the first RNA therapy to hit the mass market. Developed initially by Alnylam for familial hypercholesterolaemia, in 2013, it was licensed to The Medicines Company in Parsippany, New Jersey, with a view to using it in people with high cholesterol levels not caused by genetic conditions. In 2018, The Medicines Company sold off its other assets to focus on inclisiran exclusively.

The drug inhibits the synthesis of PCSK9, an enzyme that regulates the uptake of LDL cholesterol by liver cells, such that lowering PCSK9 expression reduces LDL cholesterol levels in the blood (high LDL levels are another risk factor for heart disease). A phase III trial of the drug in people who had experienced a heart attack or stroke showed a roughly 50% decrease in LDL cholesterol levels.

Peter Wijngaard, chief development officer at The Medicines Company, says the drug's appeal is that it needs to be given only twice a year, which guarantees six months of lower cholesterol after each physician-administered dose. The company is preparing to file an application for FDA approval later this year. In the long term, if the drug is successful in people at high risk of a cardiovascular event, it could be used in prevention and as an alternative to statins, Wijngaard adds.

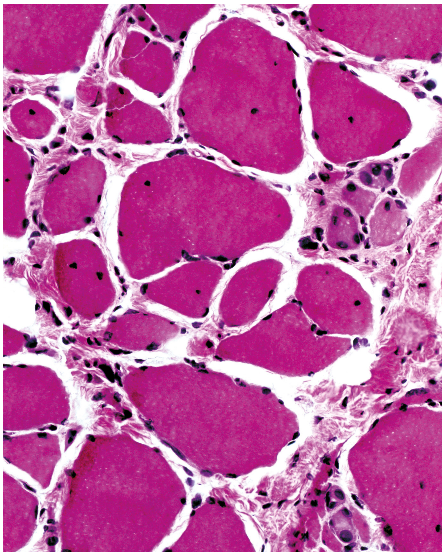
## BEYOND THE LIVER

Blocking the translation of mRNA into protein is the most straightforward and common application of RNA therapies. Indeed, catalysing mRNA degradation is the only mechanism of action of siRNA-based drugs. But ASO technology has another string to its bow.

When an mRNA is transcribed from a gene, it contains both introns and exons. Exons are stretches of mRNA that dictate the sequence of the resultant protein. Introns are stretches that are excised before protein synthesis begins, in a process called splicing. ASOs that bind to an unprocessed mRNA can influence how it is spliced into a mature mRNA.

Two drugs that alter RNA splicing have now been approved by the FDA, and, notably, neither targets the liver. One, aimed at skeletal muscle, was controversially approved for treating Duchenne muscular dystrophy (DMD). The other is RNA therapy's first blockbuster drug — a highly effective treatment for spinal muscular atrophy that operates in the central nervous system, a tissue that has become the field's second frontier.

DMD is caused by mutations in the gene that encodes the protein dystrophin, which protects muscle fibres from the damaging effects of contraction. When a mutation causes the loss of functional dystrophin,



**Light micrograph of tissue affected by Duchenne muscular dystrophy, in which some of the muscle fibres (dark purple) have been destroyed.**

BIOPHOTO ASSOCIATES/SPL

muscles progressively deteriorate and then die. Dystrophin mRNA typically comprises 79 exons. Importantly, even if an exon is missing, functional dystrophin can still be made. ASOs can be used to treat some DMD-causing mutations that render the mRNA downstream of the mutations unreadable, such that no functional protein is formed.

In the early 2000s, geneticists showed that ASOs binding to dystrophin mRNA could influence splicing in such a way that an exon containing a harmful mutation was left out of the mature mRNA (A. Aartsma-Rus *et al. Neuromusc. Disord.* **12**, S71–S77; 2002). Known as exon skipping, this resulted in the downstream mRNA no longer being unreadable and in functional dystrophin being formed.

Consequently, two groups of researchers set to work on developing exon-skipping ASOs for DMD. One team, led by geneticist Annemieke Aartsma-Rus at Leiden University Medical Center in the Netherlands, tested a compound in nearly 300 people, but it failed to reliably improve clinical outcomes.

The second team — led by researchers from Sarepta Therapeutics in Cambridge, Massachusetts — showed that its drug, eteplirsen (Exondys 51), raised dystrophin levels in recipients, but only by about 0.28%. The drug was, however, approved in a contentious FDA hearing in which family members of people with DMD gave impassioned pleas in favour of the drug and the chairperson overruled the recommendation of an expert panel. The EMA, when presented with the same data, refused to approve the application.

The field now anxiously awaits completion of an obligatory post-approval clinical trial that must show improvements in participants' symptoms. Aartsma-Rus worries that the FDA's decision might backfire: if the follow-up trial fails, regulators could hesitate to approve future drugs solely on the basis of them increasing protein levels, even when those increases are large.

> *"If I'd done nothing except Spinraza in my career, I'd be happy."*

She is working on the next generation of ASOs. "Exon skipping still has a place," she says, "but there's room for improvement." Aartsma-Rus's optimism stems from the fact that eteplirsen is an early type of ASO that doesn't contain the chemical modifications that have subsequently improved ASO efficiency. But targeting muscle tissue remains a challenge.

Another reason to be hopeful is the success of a blockbuster drug for spinal muscular atrophy (SMA), a genetic condition that affects motor neurons. It progressively diminishes a person's ability to move and, in children, can fatally disable the muscles that control breathing.

The condition occurs when a person inherits two faulty copies of the gene *SMN1*. But the human genome contains another version of this gene, *SMN2*, that is expressed at a much lower level. The protein encoded by *SMN2* can, however, substitute functionally for that encoded by *SMN1*. Nusinersen (Spinraza), an ASO developed by Ionis, binds to *SMN2* mRNA to promote the retention of an exon that increases the production of SMN2 protein to such an extent that it can compensate for the absence of SMN1. Frank Bennett, co-founder of Ionis, who leads its drug-development programme for SMA, had concerns about tackling nervous-system disorders using ASOs. Although he knew that the drugs worked well in neural tissue, he also knew that they did not cross the blood–brain barrier. To treat such a condition, ASOs would have to be given by lumbar puncture — an invasive procedure that it was hard to imagine subjecting patients to on a weekly basis.

**Logan Velchansky (front) began treatment with Spinraza soon after it was commercialized.**

IONIS PHARMACEUTICALS, INC.

But when Ionis began to test ASOs in the central nervous system (CNS) of rodents, monkeys and then humans, it discovered that the drugs stayed active for a long time. In people, nusinersen is effective when administered every 4–6 months. And trials of it were a success. Some of the children showed improvement and started to hit developmental movement milestones. Those who began the treatment before their symptoms appeared fared even better.

Bennett and his collaborator Adrian Krainer, a splicing researcher at Cold Spring Harbor Laboratory in New York, received a 2019 Breakthrough Prize for their work on the drug. Crooke says, "If I'd done nothing except Spinraza in my career, I'd be happy."

## NEXT STEPS

Such is the lasting effectiveness of ASOs in neurons that tackling CNS disorders represents a major goal for RNA-therapy researchers. Ionis has gained positive phase I/II trial results in people with Huntington's disease and similar phase I results in those with familial amyotrophic lateral sclerosis (also known as familial motor neuron disease). It has also launched ASO programmes for other neurological diseases. In April 2018, biotechnology company Biogen in Cambridge, Massachusetts, invested US$1 billion in Ionis's neuroscience programme. And a year later, Regeneron Pharmaceuticals of Tarrytown, New York, backed Alnylam's eye and CNS RNAi programme to the tune of $800 million, with Alnylam set to develop siRNA-carrying molecules analogous to GalNAc for use in those tissues. This work forms part of a dynamic ecosystem of drug development that involves at least 18 companies that are participating in ongoing clinical trials of more than 50 RNA-targeting drugs.

As well as increasingly eyeing up common disorders, ASO developers are helping to tackle extremely rare conditions. An initial promise of ASOs was that, after a suitable chemical structure had been achieved, subsequent ASO drugs based on the same structure — varying only by nucleotide sequence — would have a similar pharmacology. Each fresh sequence would have its own off-target effects, but its well-characterized basic structure would enable pharmacologists to predict tissues and cells that the drug would target, and how often a dose would need to be administered. (By contrast, for small-molecule drugs, these parameters must be determined from scratch.)

Such cumulative knowledge is useful for developing any drug, but it could be transformative for extremely rare conditions. For example, ASOs tailored to specific genetic mutations are being used by doctors at Boston Children's Hospital, Massachusetts, to treat a nine-year-old girl with the hereditary neurodegenerative condition Batten disease, and by neurologists at Columbia University Irving Medical Center in New York City to treat a 26-year-old woman with a rare form of familial amyotrophic lateral sclerosis.

Assessing drug effectiveness in trials involving only one patient is challenging — unless there are very clear effects. But they are helping to establish a framework for developing therapies for extremely rare genetic conditions.

RNA therapies stand on the cusp of a period that will define their wider adoption. There will be further cases in which RNAi- and ASO-based treatments go head-to-head, although Aartsma-Rus says that it won't be "as simple as one is always better than the other". The specifics of each condition and its treatment strategy will determine which will be the best therapeutic agent.

There is also increasing success using gene therapy, as well as the excitement that swirls around the gene-editing technology CRISPR–Cas9. Crooke sees that as a perfect example of the early wave of hope that fresh technologies attract, which isn't to say that it won't become an influential clinical tool — just that there is much work still to be done. "There will be competition for all these drugs," says Crooke. "That's the nature of these things." ■

---

**Liam Drew** *is a freelance writer based in London.*

# RNA-directed therapeutics at Ionis

**AUTHOR**
Stanley T. Crooke
Ionis Pharmaceuticals, 2855 Gazelle Ct., Carlsbad, CA 92010, Akcea Therapeutics,
22 Boston Wharf Rd, 9th Floor, Boston, MA 02210 United States

As scientists, our focus is on creating a different future. Because we are engaged in the discovery and development of new medicines, we have the extraordinary responsibility to prolong and enhance the lives of other humans. We do that by better understanding how cells perform their functions, how healthy homeostasis is maintained and what causes a disease phenotype. We then use that information, coupled to drug discovery technologies, to create medicines that have acceptable therapeutic indexes or risk–benefit ratios. Thus, as scientists, we think of history as the prologue to the future.

| | Medicines | Indication | Partner | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|---|---|---|
| Neurological | IONIS-HTT$_{Rx}$ (RG6042) | Huntington's Disease | Roche | ■ | ■ | ■ |
| | Tofersen (IONIS-SOD1$_{Rx}$) | ALS | Biogen | ■ | ■ | ■ |
| | IONIS-MAPT$_{Rx}$ | Alzheimer's Disease | Biogen | ■ | ■ | |
| | IONIS-C9$_{Rx}$ | ALS | Biogen | ■ | ■ | |
| Severe & Rare | WAYLIVRA® (volanesorsen) | FPL | Akcea | ■ | ■ | ■ |
| | AKCEA-ANGPTL3-L$_{Rx}$ | Rare Hyperlipidemias | Akcea | ■ | ■ | |
| | IONIS-GHR-L$_{Rx}$ | Acromegaly | Ionis | ■ | ■ | |
| | IONIS-PKK-L$_{Rx}$ | Hereditary Angiodema | Ionis | ■ | ■ | |
| | AKCEA-TTR-L$_{Rx}$ | ATTR | Akcea | ■ | | |
| | IONIS-TMPRSS6-L$_{Rx}$ | □-Thalassemia | Ionis | ■ | | |
| | IONIS-ENAC-2.5$_{Rx}$ | Cystic Fibrosis | Ionis | ■ | | |
| Cardiometabolic & Renal | AKCEA-APO(a)-L$_{Rx}$ | CVD | Akcea / Novartis | ■ | ■ | |
| | AKCEA–ANGPTL3-L$_{Rx}$ | NAFLD/Met. Comp. | Akcea | ■ | ■ | |
| | AKCEA-APOCIII-L$_{Rx}$ | CVD | Akcea / Novartis | ■ | ■ | |
| | IONIS-GCGR$_{Rx}$ | Diabetes | Suzhou-Ribo (China only) | ■ | ■ | |
| | IONIS-FXI$_{Rx}$ | Clotting Disorders | Bayer | ■ | ■ | |
| | IONIS-DGAT2$_{Rx}$ | NASH | Ionis | ■ | ■ | |
| | IONIS-AGT-L$_{Rx}$ | Hypertension | Ionis | ■ | ■ | |
| | IONIS-AZ4-2.5L$_{Rx}$ | CVD | AstraZeneca | ■ | | |
| | IONIS-FXI-L$_{Rx}$ | Clotting Disorders | Bayer | ■ | | |
| Cancer | IONIS-AR-2.5$_{Rx}$ | Prostate Cancer | Suzhou-Ribo (China only) | ■ | ■ | |
| | Danvatirsen | Cancer | AstraZeneca | ■ | ■ | |
| Other | IONIS-HBV$_{Rx}$ | Hepatitis B Virus Infection | GSK | ■ | ■ | |
| | IONIS-HBV-L$_{Rx}$ | Hepatitis B Virus Infection | GSK | ■ | ■ | |
| | IONIS-FB-L$_{Rx}$ | Comp.-Med. Diseases | Roche | ■ | ■ | |

**Figure 1. Drugs in development.** Current pipeline at Ionis adapted from the annual meeting presentation June 6, 2019.

Traditional drug-discovery efforts focus on designing small molecule agents (usually <500 Dalton) meant to bind to and alter the structure and function of specific proteins. It has been apparent for many decades that this process is inadequate to meet the needs of patients or the demands of industry for new products. Moreover, these inefficient drug discovery systems often fail to take advantage of the extraordinary advances in understanding the molecular mechanisms of disease. There are numerous causes for the decline in productivity of our industry. For example, although the development of antimicrobial resistance is an issue, in the developed world infectious diseases are well managed, as are many acute illnesses. This means that the industry is, to a large extent, focused on improving the therapy of chronic degenerative diseases. The treatment of chronic diseases is intrinsically a more difficult challenge for drug discoverers and developers. Increased demands from regulators before approval of new medicines, particularly the demand for high risk, enormously costly outcome studies, are a significant determinant. The evolution of monoclonal antibodies is representative of new technology that is more efficient and takes better advantage of the advances in understanding the molecular pathology of diseases that small molecule drugs are rarely able to address. But more must be done.

## AN ENTIRELY NEW CHEMICAL CLASS OF MEDICINES

Thirty years ago, armed with a blank piece of paper and US$5.2 million, Ionis set out to create a radically different drug discovery technology. We planned to design an entirely new chemical class of medicines called chemically modified oligonucleotides that would bind to their targets through an interaction never before used for drug discovery: Watson–Crick hybridization. We also planned to focus on molecules that had never been considered as drug receptors: ribonucleic acids (RNAs). We were not alone. Gilead, Hybridon and Genta were founded before or contemporaneously with

Ionis, but we stayed the course. Because of 30 years of progress, we believe that we can realize an extraordinarily exciting future. The next few pages introduce the future envisaged by Ionis.

At Ionis, we have developed and achieved regulatory approval for five first-in-class medicines, including the first RNA-targeted blockbuster, Spinraza. Currently, we have more than 40 first- or best-in-class medicines in development and expect at least 10 or more drugs to be in pivotal trials or registration[1] by the end of 2020. Because of the efficiency of RNA-targeted drug discovery as practiced at Ionis and our unique business model, we produce one medicine for every 12 Ionis employees (we refer to ourselves as Ions). That represents a large and growing competitive advantage compared to the 1,000 plus persons required for the fully integrated companies using other available drug discovery platforms. Today, we have fully validated intravenous (IV), subcutaneous (SC) and intrathecal (IT) administration. We also have encouraging data on aerosol administration and can give antisense oligonucleotides (ASOs) as enemas (**Fig. 1**). Over the next several years, we intend to grow our pipeline to approximately 60 medicines, with 10 being in pivotal trials at steady state. Because of the efficiency of the platform we can enlarge our pipeline with very modest increases in our staff. This support assures the core of Ionis never exceeds 500 or so people. Why 500? Because organizational studies have shown that about 500 is the optimal size to assure maximal innovation.

## INNOVATION THAT BRINGS MAXIMUM VALUE TO PATIENTS

In our business, the number of medicines in development matters a great deal because they provide more opportunities for success. However, our focus goes well beyond numbers. We are intent upon pursuing novel molecular targets that have the potential to revolutionize the therapeutic landscape and patient lives. Given the versatility, breadth of utility and the success rate of medicines derived from our platform, we can afford to take greater target risk. It is only by taking such an approach that transformative first-in-class medicines can be developed. We are committed to bringing the maximum value to the patients who need it the most because we feel that should be the mandate of our industry. And because we strongly believe that providing value to patients and charging for the value provided is the future of our industry. ASOs are a direct route from the gene to the patient and we fully expect that genomic advances will be continually incorporated in all our drug discovery and development activities.

In our pipeline of 2025, we expect to have fully enabled aerosol delivery of ASOs for respiratory diseases such as chronic bronchitis, commercially attractive, orally administered medicines and new chemical approaches that provide potent agents to treat diseases of the heart and skeletal muscles. Because ASOs that are members of the same chemical class have very similar properties (differing only in their sequence), we pioneered informative systems that integrate all safety data from studies in non-human primates and all randomized, placebo-controlled trials and have published the results of those analyses[2,3,4,5,6]. As new chemistries and routes of administration evolve, we will assemble and publish the results. These databases are of extraordinary value because learnings from previously developed medicines help inform our expectations to newer medicines in a particular chemical class. This contributes to the much higher success rate we see compared to more traditional approaches.

Although we are committed to bringing new therapies like Spinraza, Tegsedi and Waylivra to patients with rare diseases, our platform is also capable of producing medicines to treat more common disorders. Today, about 30% of our pipeline is committed to the treatment of rare diseases. By 2025 we expect the pipeline to be about 20% rare disease and 80% more common diseases. Importantly, we expect the performance of our medicines to continue to improve. The improved performance will be driven by advances made by the Ionis core antisense program. Today, we have medicines that can be administered subcutaneously at 10-20 mg per month or less frequently. We expect the potency of our agents to continue to increase, enabling even more cost effective and better tolerated agents. The increase in potency should support the treatment of new types of diseases such as cardiac arrhythmias and congestive heart failure.

## ON THE CUSP OF 'DESIGNER' MEDICINES

Certainly, an exciting future, but the most exciting opportunities are enabled by very recent advances in understanding the molecular mechanisms by which our medicines distribute *in vivo* and within cells and the molecular mechanisms by which they produce both pharmacodynamic and toxic effects. By exploiting this new knowledge, we believe we are on the cusp of achieving the dream of designer medicines. For the first 25 years, we focused our medicinal chemistry efforts on enhancing the interactions of ASOs with targeted RNAs. That effort has yielded what we call second generation and generation 2.5 ASOs that are substantially more potent, can be administered much less frequently by essentially all routes and are much safer than first generation ASOs. Given that we literally began with a blank piece of paper, this effort has been extraordinarily successful and resulted in our dominant intellectual position in RNA targeted therapeutics.

About five years ago, based on important new data and understanding, our focus in research to advance the performance of ASOs expanded to include attempting to improve the interactions of ASOs with target RNAs. Two simple sounding statements describe the insights that have profoundly influenced our research and the future we think is possible for the technology. The fate of phosphorothioate (PS) ASOs is defined by the proteins with which they interact. PS ASOs can significantly alter the fate of many proteins with which they interact. These observations, combined with the knowledge that as much as 99% of a dose of an ASO is not localized at the desired sites for maximal activity, resulted in the establishment of four new research directions: (i) ligand conjugated ASOs or LICA tissue targeting; (ii) the targeted intracellular distribution; (iii) the design of ASOs to take advantage of newly identified molecular mechanisms of action and; (iv) the molecular toxicology program.

Now that we have substantial knowledge about the key proteins with which ASOs interact in plasma and other *in vivo* sites, we have established an effort to conjugate ligands (LICA) to ASOs to direct a greater fraction of a dose to desired tissues and cells. The initial and, to date, most important success was achieved first at Alnylam[7] and then quickly adopted by Ionis. By conjugating N-acetylgalactosamine (GalNAc) to an ASO, it is possible to take advantage of a high capacity receptor on the surface of hepatocytes, the asialoglycoprotein receptor, to enhance productive delivery of ASOs to hepatocytes. This has consistently increased the potency of second generation ASOs for hepatic targets by approximately 30-fold. When conjugated to the generation 2.5 ASOs, which are about 10 times as potent as generation two drugs, this ligand increased potency again by about 30-fold. Based on this success, we are now concentrating efforts on identifying ligands that enhance delivery of ASOs to other tissues and cells. In collaboration with our colleagues at Astra-Zeneca we recently reported that by attaching a GLP-1 peptide to an ASO we can achieve therapeutic ASO concentrations in the beta cells of the pancreas[8]. As this ligand is optimized, we expect to create medicines that enhance the metabolic control effected by these cells. We have also made significant progress in identifying ligands that enhance delivery of ASOs to the heart, skeletal muscle and other cells. Although more work is required, we are optimistic that we will enhance the performance of ASOs in many tissues that intrinsically take up lower fractions of a dose of an ASO.

Similarly, cellular uptake and distribution of ASOs is very inefficient. Based on recent work performed primarily in our laboratory, we now understand the major pathways by which cells take up and distribute PS ASOs, the major proteins that are involved and the main sites in the cell in which ASOs are active. We also have a solid understanding of the kinetics of these processes (**Fig. 2**). Importantly, we have identified straightforward medicinal chemical modifications that can substantially alter the interactions of PS ASOs with the proteins responsible for the transport of ASOs across the plasma membrane and those that distribute ASOs inside the cell. We know that approximately 99% percent of the ASOs in the cell are in locations in which the ASO is not active, so a shift of just 1% from an inactive intracellular site to an active site should double potency[9].

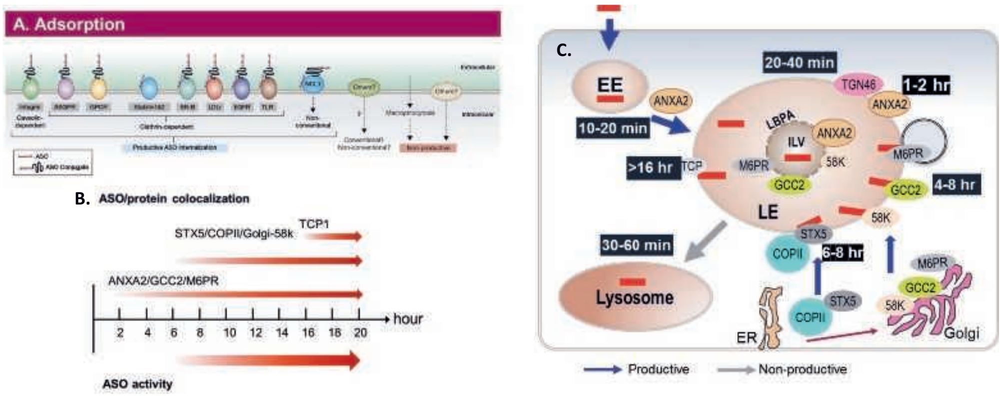


**Figure 2. The uptake and distribution pathways and correlation with ASO activity[8].** (A) Cell surface interactions. The proteins located on the cell surface that are known to interact with PSASOs and/or ligands conjugated to ASOs. The interactions with proteins that are known to enter pathways that lead to observable ASO activity in cells (productive pathways) and those that appear to result in cell uptake, but no ASO activity (non-productive) are shown. (B) The timeline of PSASO localization in key endosomal structures and with key proteins vs the ASO activity. (C) Detailed time course of PSASO trafficking in productive endosomal pathways. Adapted from (9).

The third core research theme that we expect to have a significant impact on the future of the technology is actually not a new effort. We have worked for 30 years to understand the major steps involved in the molecular mechanisms of pharmacodynamics, the kinetics of these steps and the factors that influence the effectiveness of the desired pharmacology of ASOs. We have achieved most of those goals[10]. Recent progress in this area is yielding data that potentially enhance the future value of the platform. For example, we now understand one of the major mechanisms of tolerance to ASOs and this process also likely explains why some transcripts are less effectively reduced by ASOs designed to activate RNAse H1. Perhaps, more importantly, we have expanded the repertoire of mechanisms for which we can design ASOs including those that, for the first time, selectively increase the translation of specific proteins, thus serving as 'agonist -like drugs'[11].

## OPPORTUNITIES TO DESIGN EVEN BETTER ASOS

The fourth new research direction is arguably the most important because it resolves one of the most controversial topics in the technology and provides straightforward medicinal chemical solutions to the issue. The problem is that as ASOs are screened, we occasionally encounter some compounds that are toxic. The toxicities are influenced by the sequence of the ASO and the type of 2'chemical modifications employed. We outlined a step-by-step molecular mechanism that accounts for these observations (**Fig. 2**) and a straightforward medicinal chemical approach that ablates or substantially reduces toxicities with minimal effects on potency[10]. We have expanded the database supporting this mechanism and greatly expanded the structure activity and sites of placement of modifications that ablate or reduce toxicities and enriched our understanding of the mechanism. These observations have further enhanced the efficiency of already very efficient screening methods, enhanced the overall potency of agents for development and opened new areas for medicinal chemistry.

Considering the impact of the four research themes enumerated, it is clear that we are close to realizing the dream of designer medicines: therapeutic agents that distribute in the body to perform the jobs for which they are designed, localize intracellularly to be more potent and less likely to have undesired effects, meeting a wide range of pharmacologic needs through a variety of pharmacodynamic mechanisms, and finally, avoiding the possibility of side effects other than those that might be associated with the molecular target. These are realistic opportunities that are the product of knowledge accumulated at Ionis through 30 years of persistent and unprecedented efforts in the history our industry. That, to Ionis, is an extraordinary future.

## REFERENCES

**1.** Crooke, S.T., Witztum, J.L., Bennett, C.F., & Baker, B.F. *Cell Metab.* **29** (2):501 (2019)

**2.** Crooke, S. T. *et al. Nucleic Acid Ther.* **29**, 16–32 (2019).

**3.** Crooke, S.T. *et al. Nucleic Acid Ther.* **28** (1), 10–22 (2018).

**4.** Crooke, S.T. *et al. Nucleic Acid Ther.* **27**, 121–129 (2017).

**5.** Crooke, S.T. *et al. Mol. Ther.* **24**, 1771–1782 (2016).

**6.** Rajeev, K. G. *et al. Chembiochem* **16**, 903–908 (2015).

**7.** Ämmälä, C. *et al. Sci Adv.* **4**, eaat3386 (2018).

**8.** Crooke, S.T., Wang, S., Vickers, T.A., Shen, W. & Liang, X. *Nat. Biotechnol.* **35**, 230–237 (2017).

**9.** Liang, X.H., Shen W, Sun H, Migawa M.T., Vickers T.A. & Crooke S.T. *Nat. Biotechnol.* **34**, 875–882 (2016).

**10.** Shen, W. *et al. Nat. Biotechnol.* **640**, 640–650 (2019).

**11.** Crooke, S.T. *Nucleic Acid Ther.* **27**, 70–77 (2017).

Staff members at the Allen Institute in Seattle, Washington, a non-profit research organization that includes the Allen Institute for Brain Science.

ALLEN INSTITUTE

# Brain work

## Large-scale national research projects hope to reveal the secrets of the human brain.

BY NEIL SAVAGE

Christof Koch was finishing his PhD on theoretical brain modelling in 1982 when he received a worrying telegram from his adviser, Tomaso Poggio.

Poggio, who had relocated to the United States from what was then West Germany the previous year, warned Koch, who was planning to join him, that he might struggle to find a US position. After getting his PhD from the Max Planck Institute for Biological Cybernetics in Tübingen, Koch had hoped to take up a postdoc role in Poggio's new laboratory at the Massachusetts Institute of Technology in Cambridge.

But Poggio had discovered that there was already someone in the United States making computational brain models and, Koch says, he wrote, "I'm worried, are there enough faculty positions to cover two people modelling the brain in the entire United States of America?"

That seems funny now, says Koch, because today there's no shortage of such positions; he estimates that 5,000 people in the United States do something similar. "Now every university will have two, three, four, five faculty that do nothing but theoretical neuroscience, computational neuroscience, data science in the brain," he says. And there are now many other jobs in neuroscience, at places ranging from conventional research universities and dedicated institutes to companies developing neurotechnology. "There's a variety of opportunities, vastly more than ever before," Koch says. "It's a fantastic time to study the brain."

### ON THE MAP

Many of those opportunities come from the rise of large-scale national and international projects dedicated to understanding the human brain. In the past six years, the European Union, the United States and Japan have all launched multi-year, multimillion-dollar projects to suss out the workings of the brain — how it learns, how it controls behaviour and how it goes wrong. Koch himself participates in some of those projects in his ►

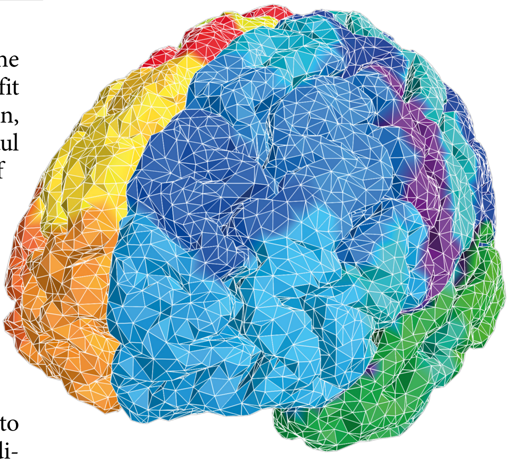
▶ role as chief scientist and president of the Allen Institute for Brain Science, a non-profit research organization in Seattle, Washington, endowed by the late Microsoft co-founder Paul Allen. The institute is building a catalogue of the various cell types in both human and mouse brains, called the Allen Brain Atlas.

Other countries — Canada, Australia, South Korea and China — are in the early stages of their own large-scale efforts. Along with the goal of describing in detail just how the brain works at various levels, from the cellular to the behavioural, the hope is that these projects will lead to new ways to treat brain diseases and mental-health conditions, as well as advance artificial-intelligence (AI) technologies. To do that, investors are providing the projects with billions of dollars in new funding, creating career opportunities for not only neuroscientists but also physicists, mathematicians, chemists, materials scientists and medical specialists, all of whom must be trained to work across disciplines.

The European Union's effort, a 10-year, roughly €1-billion (US$1.1-billion) programme called the Human Brain Project (HBP), was launched in 2013. The sprawling endeavour directly employs 500 scientists spread across universities, hospitals and research centres throughout Europe, and provides funding that supports many more. Its aim, explains scientific research director Katrin Amunts, who studies the brain at Heinrich Heine University Düsseldorf in Germany, is to develop tools for studying the brain and create an infrastructure of supercomputers and data-sharing platforms that will allow researchers to do more comprehensive studies than they could on their own. "In ten years we will not have understood the human brain. It is simply too complex and we still have to learn a lot," she says.

That does not mean, however, that the world will have to wait for some vague future date to see any benefits. For instance, Viktor Jirsa, a physicist who directs the Institute of Neuroscience Systems at Aix-Marseille University in France, has used information about how individual neurons are linked together in the brain — the 'connectome' — to develop the Virtual Epilepsy Patient. He takes readings of the brains of people experiencing epileptic seizures and uses the measurements to pinpoint more precisely the sources of those seizures. The hope is that, by targeting specific neural connections involved in a particular person's epilepsy, brain surgeons can perform more targeted treatments and improve the success rate of surgery to stop epileptic seizures beyond the current success rate of roughly 50%. Jirsa is in the middle of a randomized clinical trial to test the idea on about 350 people.

Another practical benefit Amunts hopes the HBP will yield is a visual prosthesis that could restore sight to people who are blind. This is being developed by neuroscientist Pieter Roelfsema, who leads the Vision and Cognition Group at the Netherlands Institute for Neurosciences in Amsterdam. Building a bridge between the biological understanding of the visual cortex and the engineering result of a robotic eye is the sort of outcome the HBP aims to achieve, she says. Along the way, researchers are developing better AI algorithms to decipher brain functions, and virtual robots in which to test whether the mechanisms they think they've uncovered would work in people.



**Visualization of the whole brain, simulated using real data from TheVirtualBrain framework. Colours represent different groups of connected neurons.**

THEVIRTUALBRAIN

Amunts doesn't expect the HBP to uncover all of the brain's secrets, and welcomes other national programmes. "Any project in the world cannot address all questions related to the human brain," she says.

## BUILDING CIRCUITS

With megaprojects such as the HBP, as well as other, smaller national programmes (see 'Projects around the globe') and their associated funding, neuroscience is a growth area. "It's a really great time for young people to enter the field," says Walter Koroshetz, a neurologist who directs the US National Institute of Neurological Disorders and Stroke. Although there's plenty of opportunity for researchers in biology and medicine, there's also a great need for people from other disciplines — particularly from fields that will help to develop tools to push the research forward. "The people who are going to build those are not the biologists, they're the engineers and the materials scientists and the physicists," says Koroshetz. "And the people who are going to look at neural activity and try to figure out what it means, what's the language of the brain, are going to be mathematicians and computer scientists."

In the United States, the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative launched in 2014, with a planned budget of between $300 million and $500 million a year for 12 years. Much of it is conducted under the auspices of the US National Institutes of Health. Also participating are the National Science Foundation; the Defense Advanced Research Projects Agency, which develops cutting-edge technologies for military use; and the Intelligence Advanced Research Projects Activity, which does the same for the intelligence field.

The goal during the project's first five years , says Koroshetz, was to develop tools that would allow researchers to monitor and manipulate brain circuits — the collection of neurons that work together to perform particular functions, such as coordinating muscle movement. Scientists have an understanding of the brain at the cellular level, including how genes influence the form of brain cells, how neurotransmitters and receptors work and how synapses are formed and strengthened. But they didn't have the tools to watch and understand the activity of circuits, the next level up in the hierarchy of brain function. Tools that might help include proteins that can be inserted into specific cell types to make them fluoresce when calcium flows through an ion channel, for instance,

## NATIONAL INITIATIVES

### *Projects around the globe*

- The Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) programme in Japan started in 2014 and is set to receive about US$30 million annually. The project is studying transgenic marmosets (*Callithrix jacchus*) as models for human disease.
- China plans to study the brains of macaques as a model for humans. The national China Brain Project, although much discussed, has yet to be launched by the government. Regional projects in Shanghai and Beijing are under way, however.
- The Australia Brain Initiative plans to have an emphasis on neurotechnology, with the hope of translating discoveries from brain research into commercial devices.
- South Korea launched the Korean Brain Initiative in 2017, with a focus on connectome mapping and developing new tools.
- The Latin American Brain Initiative, or LATBrain, received a boost at the end of August when researchers met in Uruguay to promote investment in the project.
- The Kavli Foundation, a non-profit organization in Los Angeles, California, funds research and coordinates the International Brain Initiative (IBI), a movement focused on helping various national projects to work together and share data. The IBI is creating an inventory of brain initiatives to list projects and funding throughout the world. **N.S.**

## EDUCATION PROGRAMMES

### *Getting into brain research*

The US and European brain projects, as well as other initiatives, offer support for young scientists interested in the research.

- The Human Brain Project (HBP) education programme offers web-based courses and in-person workshops to master's and PhD students and recent doctoral graduates in areas outside their core fields. The programme provides European Credit Transfer and Accumulation System (ECTS) credits.
- The annual HBP Student Conference on Interdisciplinary Brain Research brings young researchers from different fields together. The next one is in Pisa, Italy, in January 2020.
- The US National Institutes of Health BRAIN Initiative offers grants that provide two years of mentorship plus three years of funding for postdoctoral fellows working in research areas related to the project.
- The Institute of Electrical and Electronics Engineers in New York City hosts IEEE Brain, a programme to promote cross-disciplinary collaboration.
- The Neurophotonics Centre in Quebec City, Canada, offers a ten-day summer programme on applying optical imaging techniques to neuroscience. The next one is in June 2020.
- The BRAIN Initiative Alliance has created a website (see go.nature.com/35gulko) to help researchers to find funding from various US sources.
- The Simons Foundation, a philanthropic group in New York City, sponsors the Simons Collaboration on the Global Brain and offers neuroscience courses for scientists in various locations around the world, including China and South Africa.
- Wellcome, a research-supporting foundation in London, offers funding for research in neuroscience and mental health.
- The Allen Institute for Brain Science's Frontiers Group in Seattle, Washington, provides three years of funding to individuals or small groups with what it considers groundbreaking ideas. It also offers larger-scale, longer-term funding to teams at universities or other major institutions. **N.S.**

or targeted receptors that can turn the firing of neurons on or off in response to light or a chemical stimulus.

Armed with such tools, scientists can begin to look at what goes wrong in various disorders that don't have obvious physical signs, unlike Alzheimer's disease, in which plaques are present, or brain tumours. Diagnosis of many mental-health conditions is based on behaviour, but it could prove useful to see what's happening in the brains of people with these conditions, and that could also lead to new treatments. "There's no pathology in chronic pain, schizophrenia, depression or Tourette's syndrome, but they're clear circuit disorders," Koroshetz says. "You couldn't even approach trying to understand what was wrong with the circuit before."

## BRAINS NEEDED

To build and use those tools, researchers require a transdisciplinary mindset, says Yves De Koninck, director of the Neurophotonics Centre in Quebec, Canada. The centre brings together neuroscientists, physicists, chemists, mathematicians and engineers. They develop hardware (such as fibre-optic probes and microscopes), software to analyse and model brain data, and what De Koninck calls "wetware", such as tools that allow researchers to control living cells using light. "Brain science has to do with everything, from understanding the fundamentals of cell function all the way to how society is built, so we really need people that are able to bridge between different disciplines," De Koninck says.

De Koninck runs a transdisciplinary training programme at the Neurophotonics Centre. He is also part of the Canadian Brain Research Strategy, a collaboration of various Canadian brain-research groups. One emphasis, he says, is forming partnerships between institutions, disciplines and nations. "We can't, in the field of neuroscience, understand the complexity of the brain by continuing each to work alone," he says.

Amunts agrees that neuroscience needs researchers who have trained in a variety of fields. "Such people are very rare at this moment, so we are starting to educate them."

For instance, Heinrich Heine University, where Amunts directs the Cécile and Oskar Vogt Institute of Brain Research, offers a master's degree in translational neuroscience, and some of the courses students take as part of the degree are sponsored by the HBP. Early-career scientists who want to participate in brain research need to be willing to learn from colleagues in other disciplines, Amunts says.

And the field only seems likely to grow, brain researchers say (see 'Getting into brain research'). "It is a golden age of brain science," says Koch. "It also impacts people much more than studying the heart or the liver, because who you are depends on your brain. If your heart is faulty, you can get a heart transplant and you are still you. Who we are, how we think, our dreams, our aspirations, our very consciousness is very closely tied to the brain." ■

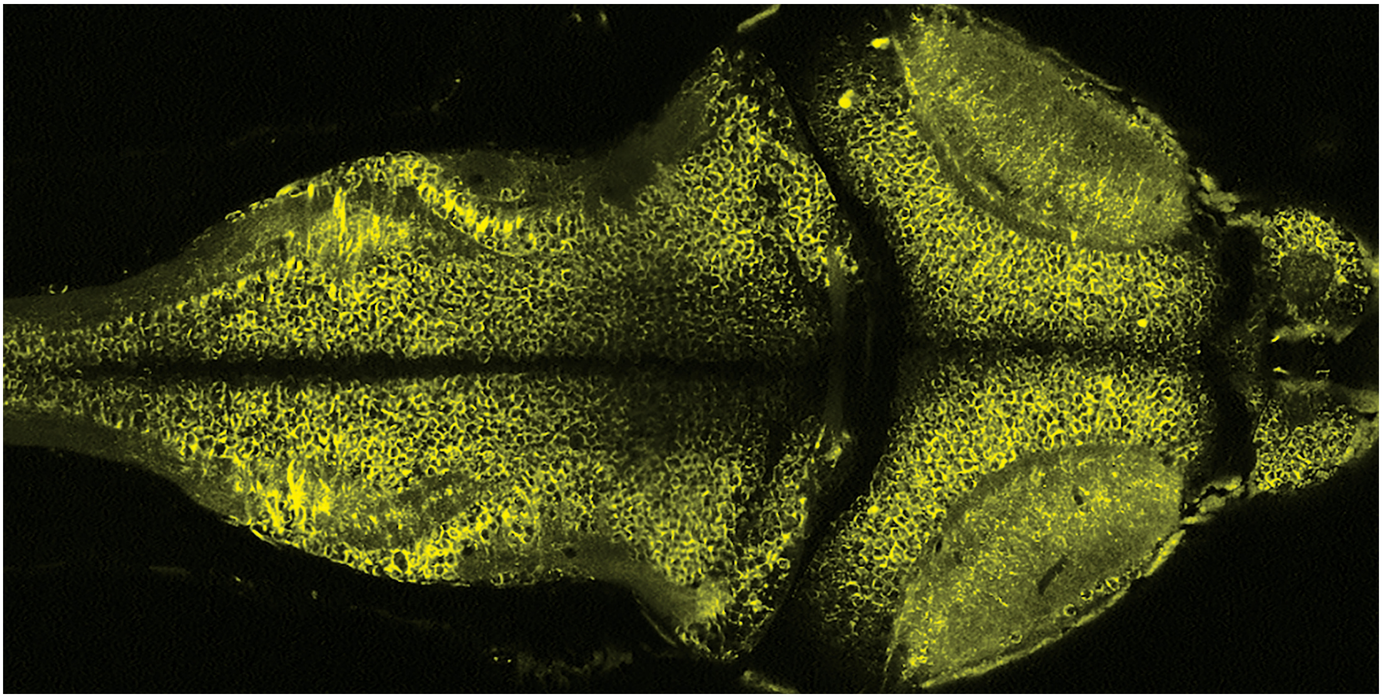
---

**Neil Savage** *is a science writer in Lowell, Massachusetts.*

TECHNOLOGY FEATURE

# GENETIC LIGHT BULBS ILLUMINATE THE BRAIN

*Genetically encoded voltage indicators change colour in real time when neurons transmit electrical information, offering unprecedented insight into neural activity.*



A. S. ABDELFATTAH *ET AL.*/*SCIENCE*

**Fluorescent proteins that react to voltage changes show signalling between cells in the brain of a zebrafish (*Danio rerio*).**

BY JYOTI MADHUSOODANAN

Douglas Storace still has the dollar bill that he triumphantly taped above his laboratory bench seven years ago, a trophy from a successful wager. His postdoctoral mentor, Larry Cohen at Yale University in New Haven, Connecticut, bet that Storace couldn't express a protein sensor of voltage changes in mice back in September 2012. Storace won.

The bill is a handy reminder that the experiments he aims to try in his new lab can work. And it's a testament to just how tricky it is to correctly express these sensors and track their signals. Storace, now an assistant professor at Florida State University in Tallahassee, plans to use these sensors, known as genetically encoded voltage indicators (GEVIs), to study how neurons in the olfactory bulb sense and react to smells.

GEVIs are voltage-sensitive, fluorescent proteins that change colour when a neuron fires or receives a signal. Because GEVIs can be targeted to individual cells and directly indicate a cell's electrical signals, researchers consider them to be the ideal probes for studying neurons. But they have proved frustratingly difficult to use. "Being able to visualize voltage changes in a cell has always been the dream," says neuroscientist Bradley Baker at the Korea Institute of Science and Technology in Seoul. "But probes that looked great often didn't behave in ways that were useful."

Early GEVIs disappointed on several levels. They were bright when a cell was resting and dimmed when the cell fired an action potential, producing signals that were tough to distinguish from the background. And they failed to concentrate in the nerve-cell membranes, where they function. But researchers are beginning to solve these issues. Some are turning to advanced fluorescent proteins or chemical dyes for better signals; others are using directed evolution and high-throughput screens to make GEVIs more sensitive to voltage changes. Meanwhile, biologists are putting these molecules through their paces. GEVIs, says neuroscientist Katalin Toth at Laval University in Quebec City, Canada, are not yet widely used, but they're getting there. "They are becoming brighter and faster — and growing in popularity," she says. "I think this is the dawn of GEVIs."

## GLIMMER OF PROMISE

When a mouse smells a banana and races towards the treat, it is the inevitable result of a well-organized orchestra of neural circuits. Researchers can tap into these pathways using patch-clamping (in which electrodes and pipettes are placed on cells to track electrical activity in a given brain region) and voltage-sensitive dyes (which can reveal overarching electrical changes).

Genetic probes are another option. In a similar way to dyes, these molecules fluoresce in response to electrical signals. But researchers

▶

▸ can use genetic tricks to limit the probes' expression to specific cells. Genetically encoded calcium indicators (GECIs), such as GCaMP proteins, are made by fusing a fluorescent protein to one that can bind to calcium. Calcium floods a nerve cell after it has fired an electrical signal, causing a change in the binding protein's shape that triggers a change in fluorescence.

But GECIs are only proxies for neural electrical activity. Although they are sensitive to action potentials, which are the basic units of neural communication, they cannot capture the smaller, sub-excitatory cues that help nerve cells to compute and integrate different kinds of information.

In 1997, Ehud Isacoff at the University of California, Berkeley, developed the first GEVI, named FlaSh, by fusing green fluorescent protein with a voltage-sensitive potassium channel[1]. Imperial College London neuroscientist Thomas Knöpfel, then at RIKEN in Wako, Japan, followed suit in 2010 by fusing a voltage-sensing phosphatase enzyme derived from *Ciona intestinalis*, a marine invertebrate, to a fluorescent protein[2]. Other designs followed, including the 2012 discovery that a random mutation in one protein made it 14-fold more sensitive to voltage changes, leading to one of the biggest early GEVI successes[3], ArcLight.

Today, there are three major classes of GEVI (see 'Flavours of fluorescence'). Probes such as ArcLight fuse voltage-sensitive protein domains (VSDs) to fluorescent proteins, whereas others such as Archer and QuasAr2 rely on fluorescent membrane-spanning bacterial proteins known as rhodopsins. Ace–mNeon represents a third group known as opsin–FRET (fluorescence resonance energy transfer) probes. These molecules combine light-sensitive opsins that are similar to rhodopsin with a second fluorescent protein to create an energy transfer — detectable as a change in fluorescent colour — when the proteins are excited. "Unlike GCaMP, where everyone was focused on one scaffold, each GEVI has its own developmental path," says neuroscientist Michael Lin at Stanford University in California.

Ideally, a GEVI will yield a bright, stable signal that consistently follows a change in voltage or action potential, and produce minimal background fluorescence. But this doesn't always happen. Unlike GCaMPs, which can fill a cell's volume, GEVIs must be localized to the cell membrane to be effective. They cannot be tested in bacteria, because it is difficult to maintain a membrane potential in these cells. And the change in fluorescence when a GEVI fires is much smaller than that seen with GECIs.

The millisecond pace of neural electrical activity is also a problem, both for GEVIs and the cameras that image them. And generating a sufficiently bright signal requires intense excitation light, which can overheat cells and cause the GEVI to bleach within minutes.

As a result, most biologists still look to GCaMPs to study fine-scale neuronal activities. "What the field would love to have is a solution like GCaMP. Calcium imaging works consistently in anyone's hands," says neuroscientist Eric Schreiter at the Howard Hughes Medical Institute's Janelia Research Campus in Ashburn, Virginia. "There are very few reports of existing GEVIs being used *in vivo*, and they're quite limited in their scope."

### BRIGHTER, FASTER … BETTER?

But that is beginning to change, thanks to directed-evolution approaches, high-throughput screening strategies and more-stable fluorescent molecules.

Schreiter and his team, for instance, removed the fluorescent portions of rhodopsin-based sensors and replaced them with a protein that binds to a synthetic dye molecule in response to voltage changes. Synthetic dyes are significantly brighter and more photo-stable than fluorescent proteins. One such probe, dubbed Voltron, produced a signal that was several-fold brighter than its parent GEVIs and lasted upwards of 15 minutes without bleaching[4].

Voltron's signal is a flare of bright light against a background of unbound dye that is also fluorescing. This 'negative' signal is much harder to spot under a microscope than a 'positive' one, where the background remains dark. In subsequent experiments, Schreiter's team discovered that three specific mutations in the rhodopsin proton-transport domain reduce the protein's fluorescence when a cell is resting and thus result in a 'reverse Voltron' that produces this kind of positive signal[5]. "It's one of the rare instances in my career where trying something rational actually worked on a protein," says neuroscientist Ahmed Abdelfattah, a postdoctoral researcher in Schreiter's lab.

These mutations could also help to tweak the bright-on-bright signals from other rhodopsin-based GEVIs, says neuroscientist Yuki Bando at Hamamatsu University in Japan.

Other researchers have made GEVIs that use red fluorescent proteins instead of green ones, because red light can penetrate deeper into tissues and causes less cellular damage. Recent examples include VARNAM (voltage-activated red neuronal-activity monitor), which blends a red fluorescent protein named mRuby with an opsin-based probe; nirButterfly, a variant of the FRET-based GEVI called Butterfly that swaps the paired fluorescent proteins with bacterial near-infrared proteins; and Ilmol, which uses one of the brightest red fluorescent proteins available to produce a signal three times stronger than that of FlicR1, the VSD probe on which it is based[6–8].

Expanding the GEVI spectrum helps researchers to combine probes or techniques in the same study. For example, chemical biologist Adam Cohen at Harvard University in Cambridge, Massachusetts, and his colleagues developed QuasAr3, a near-infrared probe that surpasses its predecessors in terms of signal-to-noise ratio, membrane-specific expression and other properties. The team combined QuasAr3 with optogenetics — the use of different wavelengths of light to control neuronal activity — to study voltage changes correlated with behaviour and movement in mice[9].

"Some of the largest improvements in opsin-based probes have been in membrane localization," Lin says. "That alone has been very useful. Others such as nirButterfly and VARNAM also show improvements in brightness and responsiveness."

Lin's own work has focused on a series of GEVIs dubbed ASAPs for their fast responses. The latest iteration, ASAP3, has a signal that is significantly stronger than its predecessor, ASAP2 (ref. 10).

### SPEEDIER SCREENS

Many of these protein improvements stem from directed-evolution techniques, in which proteins are randomly mutated and improved versions selected over multiple cycles. But it is one thing to create a GEVI, and another thing to test it. When it comes to GEVIs, says protein engineer Robert Campbell at the University of Alberta in Edmonton, Canada, screening remains a "bottleneck".

Yale University neuroscientist Vincent Pieribone's team, which created VARNAM, uses a 96-well-plate set-up in which a field electrode moves from well to well to excite cells that carry GEVIs harbouring different mutations, linkers or fluorescent proteins. The system lets researchers quickly study each GEVI's response to voltage changes. In another approach, Edward Boyden at the Massachusetts Institute of Technology in Cambridge enriched for brighter rhodopsin-based probes using a microscopy-guided, robotic process[11].

Lin's group turned to a classic genetic technique — electroporation — to quickly screen its probes. In electroporation, a quick pulse of electricity reduces the resting membrane potential of cells to zero, creating temporary holes in the cell membrane so that DNA can enter. But because electroporation creates a defined voltage change, it can also be correlated to a probe's fluorescence signal. "It's a very simple idea," Lin says, "But it hadn't been used before to screen fluorescent probes."

These high-throughput methods help researchers to screen thousands of GEVI variants in a matter of hours. "Perhaps a major push to apply such screening systems to the right templates" is all that's needed to propel GEVI technology forward — and into more widespread use, Campbell suggests.

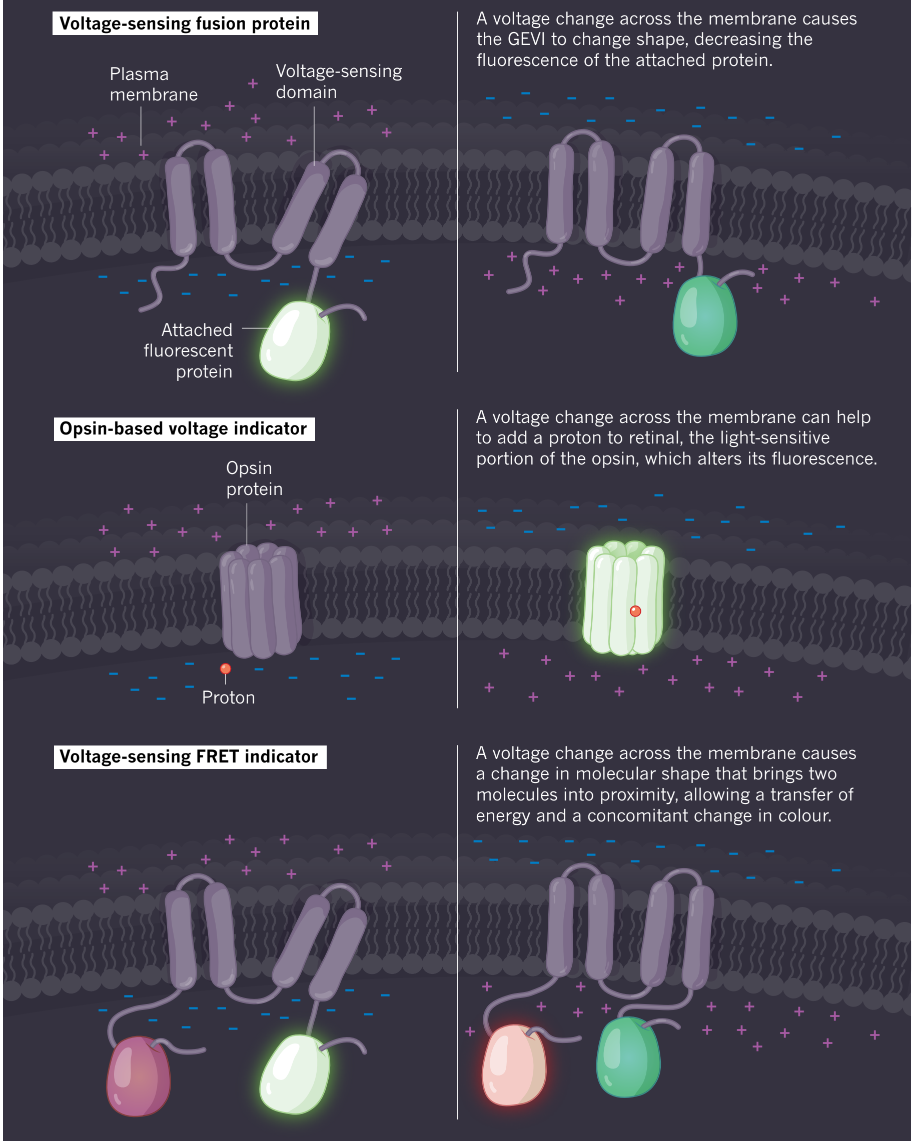
### BRIGHT APPLICATIONS

GEVI developers usually work with collaborators to test their probes in flies, mice or cultured human cells. Although independent researchers can order and use any GEVI from the non-profit repository Addgene, based in Watertown, Massachusetts, there have been few published reports from external users who have applied GEVIs to their experiments.

In part that could be due to instrumentation, Toth says. Her lab works with the ASAP

## FLAVOURS OF FLUORESCENCE

Scientists have built different types of genetically encoded voltage indicator (GEVI). One major category (top) uses a membrane-bound portion of a voltage-sensing protein, such as a sodium channel, fused to one or more fluorescent proteins. Another category (middle) uses an opsin protein, such as a microbial rhodopsin, a membrane channel that directly changes its fluorescent properties in response to an electric field. Other categories rely on fluorescence resonance energy transfer (FRET) to produce signals (bottom).



**Voltage-sensing fusion protein**

A voltage change across the membrane causes the GEVI to change shape, decreasing the fluorescence of the attached protein.

**Opsin-based voltage indicator**

A voltage change across the membrane can help to add a proton to retinal, the light-sensitive portion of the opsin, which alters its fluorescence.

**Voltage-sensing FRET indicator**

A voltage change across the membrane causes a change in molecular shape that brings two molecules into proximity, allowing a transfer of energy and a concomitant change in colour.

probes, and uses random-access two-photon microscopy to capture signals. But these systems can prove expensive and thus difficult to access for many researchers, she says.

Another hurdle has been the difficulty researchers have in directly comparing GEVI performance. Most labs report GEVI performance using a few standard metrics, which can help users to decide how different probes stack up against each other, Lin says. But until this year, only one study had compared a range of GEVIs in parallel using the same experimental conditions[12]. "Each GEVI has very specific characteristic properties," says Bando, who led that research. "But nobody had compared the indicators with the same experimental conditions."

Bando and his colleagues compared eight GEVIs in cultured neurons and mice, using both widefield and two-photon imaging. The team tested the proteins for their ability to track action potentials, synaptic input, photobleaching and other properties. Some probes, they found, emitted a dim baseline fluorescence and thus needed a very-high-powered excitation laser, which could overheat and damage cells. Others produced fast, reliable signals under conventional imaging, but failed with two-photon microscopy, which can visualize deeper brain regions *in vivo*. Still others produced strong, but short-lived signals. Overall, they found that "no indicators could detect both action potential and synaptic inputs *in vivo*", Bando says. And, "only ArcLight worked with two-photon imaging *in vivo*".

But ArcLight's fluorescent responses are slow — too slow to track a neuron's action potential, which lasts only about a millisecond. Nonetheless, this GEVI's consistency and clear signal led Bando to focus on optimizing ArcLight for his own studies. When selecting a voltage indicator, consider the purpose of the experiment, he suggests. Archer, QuasAr and Ace2N–mNeon are ideal for one-photon (that is, conventional) imaging of cultured cells or brain slices. To study deeper brain regions in live animals using multiphoton approaches, ArcLight might be a researcher's best choice for now.

Baker concurs that for GEVI novices, ArcLight is the easiest option. Probes can fail for several reasons, he explains, including incompatible cameras, photodamage, and poor or aberrant protein expression. "So many things can go wrong, and you need a probe that gives you the confidence that the imaging works," he says. "It might not be the right probe for you, but if you don't see a signal with ArcLight, you will not see a signal with whatever other GEVI you try."

Indeed, ArcLight's sluggishness might be the reason it works so well, Storace says, because it ends up integrating input from various neurons into a single signal that is easy to distinguish from background noise. "I strongly feel the reason it's better is because it's slower," he says.

The brain region being studied is also a factor. Storace focuses on the olfactory system, which is physically organized into bulbs, each of which responds to a single odorant. Applying GEVIs to this region reveals useful data, he says, because even if the signals from individual cells are indistinguishable, the population-level data can be revealing. "It's a useful strategy in the olfactory system, but I'm not sure it's easily translated to other brain areas such as the cortex," he says.

Indeed, as researchers continue refining and exploring these probes, fresh strategies are likely to emerge, as will insights into their strengths and weaknesses. Different probes might well be ideal for different questions, Storace says. "In about four or five years, we'll have a better idea of how easy it is to use GEVIs." ■

---

**Jyoti Madhusoodanan** *is a science writer in Portland, Oregon.*

1. Siegel, M. S. & Isacoff, E. Y. *Neuron* **19**, 735–741 (1997).
2. Akemann, W., Mutoh, H., Perron, A., Rossier, J. & Knöpfel, T. *Nature Methods* **7**, 643–649 (2010).
3. Jin, L. *et al. Neuron* **75**, 779–785 (2012).
4. Abdelfattah, A. S. *et al. Science* **365**, 699–704 (2019).
5. Abdelfattah, A. S. *et al.* Preprint at bioRxiv https://doi.org/10.1101/690925 (2019).
6. Kannan, M. *et al. Nature Methods* **15**, 1108–1116 (2018).
7. Monakhov, M. V. *et al.* Preprint at bioRxiv https://doi.org/10.1101/536359 (2019).
8. Yi, B., Kang, B. E., Lee, S., Braubach, S. & Baker, B. J. *Sci. Rep.* **8**, 15199 (2018).
9. Adam, Y. *et al. Nature* **569**, 413–417 (2019).
10. Chavarha, M. *et al.* Preprint at bioRxiv https://doi.org/10.1101/445064 (2018).
11. Piatkevich, K. D. *et al. Nature Chem. Biol.* **14**, 352–360 (2018).
12. Bando, Y., Sakamoto, M., Kim, S., Ayzenshtat, I. & Yuste, R. *Cell Rep.* **26**, 802–813 (2019).

# CAREERS

**NEGATIVE RESULTS** Publishing them improves science **go.nature.com/negative**

**IMPOSTER SYNDROME** How one researcher overcame it **go.nature.com/imposter**

**PARENTAL LEAVE** Ways to share it **go.nature.com/parent**

ALLSTAR PICTURE LIBRARY/ALAMY STOCK PHOTO

**Actor Josh Brolin as Llewelyn Moss in the 2007 film adaptation of *No Country for Old Men*, based on the novel by Cormac McCarthy.**

## COLUMN

# Tips from a Pulitzer prizewinner

**Van Savage** and **Pamela Yeh** share advice from US novelist Cormac McCarthy.

For the past two decades, Cormac McCarthy — whose ten novels include *The Road*, *No Country for Old Men* and *Blood Meridian* — has provided extensive editing to numerous faculty members and postdocs at the Santa Fe Institute (SFI) in New Mexico. He has helped to edit works by scientists such as Harvard University's first tenured female theoretical physicist, Lisa Randall, and physicist Geoffrey West, who authored the popular-science book *Scale*. Van Savage, a theoretical biologist and ecologist, first met McCarthy in 2000, and they overlapped at the SFI for about four years while Savage was a graduate student and then a postdoc. Savage has received invaluable editing advice from McCarthy on several science papers published over the past 20 years. While on sabbatical at the SFI during the winter of 2018, Savage had lively weekly lunches with McCarthy. They worked to condense McCarthy's advice to its most essential points so that it could be shared with everyone. These pieces of advice were combined with thoughts from evolutionary biologist Pamela Yeh and are presented here. McCarthy's most important tip is to keep it simple while telling a coherent, compelling story. The following are more of McCarthy's words of wisdom, as told by Savage and Yeh.

Use minimalism to achieve clarity. While you are writing, ask yourself: is it possible to preserve my original message without that punctuation mark, that word, that sentence, that paragraph or that section? Remove extra words or commas whenever you can.

Decide on your paper's theme and two or three points you want every reader to remember. This theme and these points form the single thread that runs through your piece. The words, sentences, paragraphs and sections are the needlework that holds it together. If something isn't needed to help the reader to understand the main theme, omit it.

Limit each paragraph to a single message. A single sentence can be a paragraph. Each paragraph should explore that message by first asking a question and then progressing to an idea, and sometimes to an answer. It's also perfectly fine to raise questions in a paragraph and leave them unanswered.

Keep sentences short, simply constructed and direct. Concise, clear sentences work well for scientific explanations. Minimize clauses, compound sentences and transition words — such as 'however' or 'thus' — so that the reader can focus on the main message.

Don't slow the reader down. Avoid ▶

▸ footnotes because they break the flow of thoughts and send your eyes darting back and forth while your hands are turning pages or clicking on links. Try to avoid jargon, buzzwords or overly technical language. And don't use the same word repeatedly — it's boring.

Don't over-elaborate. Only use an adjective if it's relevant. Your paper is not a dialogue with the readers' potential questions, so don't go overboard anticipating them. Don't say the same thing in three different ways in any single section. Don't say both 'elucidate' and 'elaborate'. Just choose one, or you risk that your readers will give up.

And don't worry too much about readers who want to find a way to argue about every tangential point and list all possible qualifications for every statement. Just enjoy writing.

With regard to grammar, spoken language and common sense are generally better guides for a first draft than rule books. It's more important to be understood than it is to form a grammatically perfect sentence.

Commas denote a pause in speaking. The phrase "In contrast" at the start of a sentence needs a comma to emphasize that the sentence is distinguished from the previous one, not to distinguish the first two words of the sentence from the rest of the sentence. Speak the sentence aloud to find pauses.

Dashes should emphasize the clauses you consider most important — without using bold or italics — and not only for defining terms. (Parentheses can present clauses more quietly and gently than commas.) Don't lean on semicolons as a crutch to join loosely linked ideas. This only encourages bad writing. You can occasionally use contractions such as isn't, don't, it's and shouldn't. Don't be overly formal. And don't use exclamation marks to call attention to the significance of a point. You could say 'surprisingly' or 'intriguingly' instead, but don't overdo it. Use these words only once or twice per paper.

### TONE IS IMPORTANT

Inject questions and less-formal language to break up tone and maintain a friendly feeling. Colloquial expressions can be good for this, but they shouldn't be too narrowly tied to a region. Similarly, use a personal tone because it can help to engage a reader. Impersonal, passive text doesn't fool anyone into thinking you're being objective: "Earth is the centre of this Solar System" isn't any more objective or factual than "We are at the centre of our Solar System."

Choose concrete language and examples. If you must talk about arbitrary colours of an abstract sphere, it's more gripping to speak of this sphere as a red balloon or a blue billiard ball.

Avoid placing equations in the middle of sentences. Mathematics is not the same as English, and we shouldn't pretend it is. To separate equations from text, you can use line breaks, white space, supplementary sections, intuitive notation and clear explanations of how to translate from assumptions to equations and back to results.

When you think you're done, read your work aloud to yourself or a friend. Find a good editor you can trust and who will spend real time and thought on your work. Try to make life as easy as possible for your editing friends. Number pages and double space.

After all this, send your work to the journal editors. Try not to think about the paper until the reviewers and editors come back with their own perspectives. When this happens, it's often useful to heed Rudyard Kipling's advice: "Trust yourself when all men doubt you, but make allowance for their doubting too." Change text where useful, and where not, politely explain why you're keeping your original formulation.

And don't rant to editors about the Oxford comma, the correct usage of 'significantly' or the choice of 'that' versus 'which'. Journals set their own rules for style and sections. You won't get exceptions.

Finally, try to write the best version of your paper: the one that you like. You can't please an anonymous reader, but you should be able to please yourself. Your paper — you hope — is for posterity. Remember how you first read the papers that inspired you while you enjoy the process of writing your own.

When you make your writing more lively and easier to understand, people will want to invest their time in reading your work. And whether we are junior scientists or world-famous novelists, that's what we all want, isn't it? ■

**Van Savage** *is a theoretical biologist and ecologist at the University of California, Los Angeles, and at the Santa Fe Institute, New Mexico, USA.* **Pamela Yeh** *is an evolutionary biologist at the University of California, Los Angeles, and at the Santa Fe Institute, New Mexico, USA.*

---

## COLUMN

# Use poetry to share your science

### **Sam Illingworth** explains how poetry can communicate and celebrate research.

In reading Vachel Lindsay's poem 'The Horrid Voice of Science' (Box), it is easy to see why many people perceive poetry and science to be at odds with one another: the former searches for beauty in the enigmatic and the magical, and the latter seeks to explain this mystery. As a scientist who uses poetry in my research, I feel that these two disciplines are complementary, and we should embrace poetry as a tool for communicating our research and developing new understandings of our fields.

One of poetry's most enduring qualities is its ability to consider a topic from a new point of view; see, for example, 'A Brief Reflection on Accuracy' by the Czech immunologist and poet Miroslav Holub. This makes poetry an effective medium for disseminating scientific concepts to non-specialists. Initiatives such as the Sciku Project (which presents scientific discoveries and ideas in a haiku format), the Universe in Verse (an annual poetic celebration of science) and my own blog the Poetry of Science (in which I write a weekly poem based on published research) introduce new scientific ideas in a digestible form that is free of the jargon and technical language that can turn non-specialists away from even the most lucid abstracts. These projects do not aim to fully explain detailed scientific results through laboured rhyme. Instead, they introduce the reader to new research, encourage them to find out more about a topic and help to make science more accessible to a wider audience.

These science-communication initiatives all demonstrate how poetry can be used to communicate science to non-scientists. But they are limited in the direction of their flow of knowledge. Here, scientists are writing poetry for non-scientists — and are not necessarily concerned with how such an audience could influence their research. Poetry is arguably even more effective in developing dialogues between scientists and the broader public; for example, the Experimental Words project, funded by Arts Council England, has brought together poets and scientists to create works of art that explore the spaces between the two disciplines.

These dialogues can present researchers with insights into the direction and governance of future research. This sort of approach has been used to discuss topical issues, such

GRANGER HISTORICAL PICTURE ARCHIVE/ALAMY

The poet Vachel Lindsay reads with one of his relatives.

as climate change and disaster resilience, with often under-served audiences (for example, religious communities, older people and those living with mental-health conditions). The experiences and needs expressed in this poetry can then be used to shape policy in these areas.

Many scientists might find this conversation-through-poetry approach either too resource intensive or outside their comfort zone — or both. But poetry can benefit researchers in another way.

There are times when scientists work on a problem that seems to be unsolvable. We stop thinking about it and focus on something else, only to return to the issue at a later date. Making time for this 'incubation period' has been shown to foster creativity in problem solving, and a study published last year highlights the idea that poetry might effectively target this incubation. In other words, if you find yourself stuck on a particular problem, leave your experiment, close your laptop, stop taking your field measurements — and try writing a poem about it instead. Doing so will help you to consider the problem from a new angle, and will likely lead you to fresh insights.

### POETRY

### Science versus the arts

*"There's machinery in the butterfly;*
*There's a mainspring to the bee;*
*There's hydraulics to a daisy,*
*And contraptions to a tree.*

*"If we could see the birdie*
*That makes the chirping sound*
*With x-ray, scientific eyes,*
*We could see the wheels go round."*

*And I hope all men*
*Who think like this*
*Will soon lie*
*Underground.*

'The Horrid Voice of Science'
Vachel Lindsay (1919)

## HOW TO GET STARTED

**Read some poetry.** Just as you'd start any successful research project with an in-depth literature review to assess the field, begin by reading a wide variety of verse. This will help you to find your own poetic voice. There are many excellent examples of this art, both online and in print. Check out the magazines *Magma*, *The Rialto*, *Rattle* and *The Poetry Review*.

**Consider using a specific format.** Introducing structure to your poetry can help to reduce the anxiety of where to begin. The website Shadow Poetry offers a wide selection of poetic forms, along with examples that might provide you with scaffolding to build your poems on.

**Edit, edit, edit.** It is very unlikely that your first draft will represent the best form of your work. After you have written your poem, give it time to breathe, and then go back and remove any word that does not earn its keep.

**Think about a title.** Naming your poem can be a difficult task. The best advice I can offer comes from the excellent 2017 book *How to be a Poet* by Jo Bell and Jane Commane, which recommends that you name the work after it has taken shape. Then consider how the title interacts with the final line. Use these as a pair of bookends to frame the poem as a whole work.

**Share your poetry.** Even if this is only with a close friend or colleague, sharing your poem is a helpful way to get feedback, showcase your ideas and express your creativity. ■

---

**Sam Illingworth** *is a senior lecturer in science communication at Manchester Metropolitan University, UK.*

# THE WIND KNOWS ALL

*Defence mechanism.*

**BY BETH CATO**

The wind knows all. Even as it roars, it listens. It understands each blade of grass, each grain of dust, the very burden of its soil. Above everything else, the wind wishes its world to survive.

It knows the girl stands ready, eyes squinting as the Lady's shuttlecraft lands amid what was once the town square.

The girl is the last to remain of this aborted colony. Humans arrived here just 15 years earlier, determined to tame this far-flung world. She was the first to be born, to take her first breath on this planet her kind dubbed Tantalus. This changed her in ways left unapparent to her until recent months.

The Lady disembarks with a contingent of soldiers.

The girl meets them. She is lanky and brown as the skies.

“Maribel Rodriguez.” The Lady's voice is enhanced by her mask, which also guards her against dust pneumonia. “You've been evicted from Tantalus for your own safety. Please come with us.”

“No.” The wind lowers its voice so that hers may be heard. She's scared. She wishes her loved ones could stand beside her.

“Your parents are dead. The colony's been obliterated.” Through the face plate, the Lady's face is kind. “It's a miracle you haven't come down with pneumonia already. Doctors will examine you on the orbital.”

The Lady's orbital, the Lady's planet; as if she could own such a thing. Her corporation established the self-sufficient agricultural colony as they explored how best the planet might be utilized.

“Don't talk about miracles.” The girl once pleaded to keep her family alive. Now she understands that although the wind knows all, it doesn't control all — native spores included.

Even the wind is surprised the girl still lives.

The Lady motions to her soldiers. One fires a tranquilizer dart. The wind drops it to the ground.

“How —” the Lady begins.

“You ordered us to till wide swaths of land,” says the girl. The words aren't from her mind; this annoys her. “Our administrators argued that such practices had disastrous consequences back on Earth. You ordered the auto-ploughs anyway.”

“With the current war on Earth —”

“You knew that the loss of deep-rooted native grasses would create a ‘dust bowl’ that would probably make people sick.”

“I didn't *want* people here to get ill.” The Lady looks stunned by the accusation. “But it's true, I did encourage aggressive production goals so we could divide the population and initiate the mining operation. Earth needs those minerals to —”

ILLUSTRATION BY JACEY

“We started dying and you didn't care.”

“Of course I cared.”

“Not enough to help us, and the administrators' logs were destroyed upon their deaths.”

The wind feels the tremor of the Lady's heartbeat against her suit. “You have no proof of that.” She motions to her soldiers.

Their commander replies over a private channel. “Ma'am, she's not a direct threat, and a child besides. By intergalactic bylaws —”

“She's talking about things she can't understand. She needs treatment. We must get her off world by force.”

The girl almost breaks into hysterical giggles at the eavesdropped conversation, and instead fights to contain a hacking cough.

“The planet wants you gone. You cannot blast and mine this world as you planned.”

The Lady cocks her head. “How could you — wait, you're speaking for the entire planet? You're delusional.”

The girl thought she'd been delusional, too, when the wind first started to talk to her, when it first included her dead mother's voice, then so many others. “Tantalus isn't Earth. You can't regard it or abuse it in the same way.” These words emerge from the girl's lips, not her mind, but she fully agrees with the sentiment.

The wind chooses that moment to whip up with a hurricane's bluster. The town site vanishes within a vicious whirl of brown.

“We're not leaving without her,” the Lady yells. “Something strange is going —”

“This spontaneous storm is approximately 100 miles in diameter, centred on us, and rising in strength. Our primary duty is your well-being. We must evacuate.” Ignoring the Lady's protests, the soldiers drag her away.

The wind has some understanding of the human concepts of ‘mercy’ and ‘justice’. It ignores the soldiers, and lifts up the Lady's mask far enough to deposit spores for her next inhalation. Time will do the rest.

The shuttle rises. The wind lowers its crescendo. The girl rests on all fours, coughing. The worst of her pneumonia passed in recent weeks. She'll survive what no one else did. Sometimes, she regrets that.

“You should've let me control all of my words.” She prefers to talk out loud. “This hasn't ended, they'll soon send more shuttles, more machines —”

“You were scared.” The voice of the planet is carried by the wind, and echoed by the combined whispers of her parents and neighbours. They've melded with the planet they died on, but despite her unusual connection with this world, the girl has not.

As her gaze traces the distant speck of the shuttle, she swallows down her childhood yearnings to see Earth. She grieves and aches for the company of real people, not these blurred-together ghosts, but she needs to stay. Tantalus is her home. She'll save it.

“I'm human. It's normal to feel scared,” she says to the wind. “But I need to manage on my own. I might need to negotiate aboard the orbital some day, you know.”

The wind knows this possibility, and of fear. Humanity has taught this world what it means to be scared. The girl senses the wind's mood, and her fingers stroke the breeze, as if to calm it.

Together, they watch the shuttle fade away, vigilant for whatever may come. ■

**Beth Cato** *resides in Arizona. She's the author of the* Clockwork Dagger *duology and the* Blood of Earth *trilogy with Harper Voyager. Her website is BethCato.com.*